INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



FISICA VIBRATORIA DEL PIANO

Marzo 1979

170 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 30
Marzo 1979

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Francisco Pérez: *La producción mundial de carbón;* Miguel Viñas: *La formación de membranas celulares;* Amando García: *La óptica de ondas guiadas;* Miquel Gich: *La estructura profunda de los continentes;* Monique Robert: *Las trayectorias de los movimientos bruscos del ojo;* Andrés Lewin: *Vibraciones acopladas de las cuerdas del piano;* Joandomènec Ros: *La cabeza del cachalote;* Luis Bou: *Juegos matemáticos* y *Ciencia y sociedad;* José E. Myro y Alicia Menéndez: *Taller y laboratorio.*

Libros:

Philip Morrison, Luis Alonso, Miguel Fisac y Alfredo Carrato



LA PORTADA

El dibujo de la portada representa las cuerdas de un piano Steinway modelo B que corresponde a las notas del teclado. Cada tecla pone en movimiento un grupo de tres cuerdas que no están exactamente afinadas a la misma frecuencia. Este "desafinamiento" contribuye a la calidad del sonido del tono del piano (véase "Vibraciones acopladas de las cuerdas del piano", por Gabriel Weinreich, página 78). Al pulsar una tecla de piano, un martillo recubierto de fieltro percute el grupo mismo correspondiente de tres cuerdas. Al mismo tiempo, el bloque de fieltro blando, denominado apagador, se levanta alejándose de las cuerdas para que éstas puedan vibrar libremente. El sonido producido decae o se atenúa rápidamente al principio, y luego más lentamente. La desafinación de las cuerdas contribuye a la resonancia.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 1870 pesetas
Extranjero:
Un año (12 números): 36 U.S. $
Ejemplar atrasado ordinario:
190 pesetas
Ejemplar atrasado extraordinario:
270 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovín
Avda. de Moratalaz, 137; Madrid-30
Tel. 439 41 50 ext. 468
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Dibujo de la portada de Ted Lodigensky

Página	Fuente	Página	Fuente
9	Union Pacific Railroad Co.	68-76	Lorelle M. Raboni
10-18	Walken Graphics	79-86	Dan Todd
21	David M. Knipe, Massachusetts Institute of Technology	89-90	Tom Prentiss
		91	Ilil Arbel
22-35	George V. Kelvin	92-93	Tom Prentiss
38	Floyd C. Clark, Instituto de Tecnología de California	94-96	Ilil Arbel
		98	Malcolm R. Clarke, Marine Biological Association of the United Kingdom
39	Amnon Yariv y Pochi A. Yeh, Cal. Tech., y Alfred Y. Cho, Bell Laboratories	103-112	David Lurié, J. Wagensberg y A. Sellés
40-46	Gabor Kiss	114-119	Adolph E. Brotman
51-62	Allen Beechel	121-122	James Bell
66	John Goeller, New York Eye and Ear Infirmary	123-125	Michael Goodman

Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona.
Printed in Spain - Impreso en España

Cartas

Señor Director:

El principal motivo de esta carta es señalar que existen pruebas recientes que permiten establecer una relación entre los microfósiles descritos en el artículo del doctor Schopf ("La evolución de las células primitivas", de J. William Schopf; INVESTIGACIÓN Y CIENCIA, noviembre 1978) y las microesferas de proteinoides de las que nos habla el artículo del doctor Dickerson en el mismo número de la revista ("La evolución química y el origen de la vida", de Richard E. Dickerson; INVESTIGACIÓN Y CIENCIA, noviembre 1978). Estas nuevas pruebas han sido aportadas por una publicación reciente de Susan Francis, Lynn Margulis y Elso S. Barghoorn en *Precambrian Research*, volumen 6, páginas 65-100, 1978. Los autores han utilizado un nuevo método de fosilización artificial desarrollado por Elso Barghoorn; lo han aplicado a microesferas de proteinoides, a algas y a esporas de hongos. Sus morfologías respectivas son prácticamente imposibles de distinguir. Aunque ya se habían publicado sugerencias semejantes en torno a las microesferas no fosilizadas en las ediciones de 1972 y de 1977 de *Molecular Evolution and the Origin of Life*, de Sidney W. Fox y K. Dose (Marcel Dekker Inc., 1977), este tipo de comparaciones sirvieron de base, por ejemplo, para este comentario del profesor John Keosian en el texto de *The Origin of Life and Evolutionary Biochemistry*, de Dose et al. (Plenum Press, 1974): "Desde el punto de vista estructural, los microfósiles más antiguos se parecen más a aglutinados de un determinado número de microesferas de Fox que a algas primitivas."

Para emitir un juicio definitivo acerca de la posibilidad de que los microfósiles estén formados por microesferas de proteinoides, o cualquier otro tipo de organismo vivo, es necesario disponer de un número suficiente de fotografías, junto con datos relativos a la composición química y las actividades bioquímicas, y asegurarse de la consonancia de las observaciones con un esquema general más amplio. Como sugiere Dickerson en uno de los pasajes de su extenso artículo, las gotitas de coacervatos no nos dan una explicación de cómo se formaron los primeros polímeros biológicos en un tiempo en que no habían organismos vivos capaces de producirlos. (La protamina procede de organismos evolucionados.) Sin embargo, la simulación geoquímica que da lugar a las microesferas de proteinoides no puede ser atacada desde el mismo punto de vista.

Quisiera puntualizar algunas cuestiones. En primer lugar, el montaje escénico según el cual sería *necesaria* la formación de aminoácidos *en* el océano y su posterior calentamiento sobre un cono de escoria volcánica es idea de otros autores, no mía.

Es un requisito erróneo que he intentado corregir desde 1965. En cuanto al paso siguiente, en el que se simula la formación de protocélulas a partir de proteínas termales y agua, me pregunto de dónde sacó el doctor Dickerson la especificación de unas condiciones de temperatura de 130-180 grados Celsius. Hemos informado, en repetidas ocasiones, que la formación de las microesferas se realiza a temperatura ambiente. Es más, sería imposible utilizar una temperatura de 130 grados Celsius, ya que el agua, al hervir, abandonaría el material de vidrio no tapado que, insistimos, es preciso usar para poder simular adecuadamente las superficies lunares o planetarias. En relación con esta cuestión, digamos que Ilya Prigogine fue de los primeros en hacer hincapié, en 1955, en la necesidad de estudiar los procesos biológicos y otros fenómenos naturales en sistemas *abiertos* que no hayan alcanzado el equilibrio.

En el texto encuentro una omisión que resulta crucial en cualquier revisión general de la evolución química y es el *principio*, ampliamente aceptado, *de la autoordenación de los aminoácidos*. A nuestro parecer, este principio constituye uno de los dos "eslabones perdidos" que faltaban para poder unir la cadena de los acontecimientos evolutivos desde el estadio de elementos hasta el de la primera protocélula reproductora. He descrito este principio en el capítulo titulado "El origen y la naturaleza de la protocélula" de la obra *The Nature of Life* editada por W. Heidcamp (University Park Press, 1978), que es una recopilación de los trabajos presentados en el Conferencia Nobel sobre la Naturaleza de la Vida, celebrada en el Gustavus Adolphus College en octubre de 1977.

SIDNEY W. FOX

Universidad de Miami
Coral Gables, Florida.

Señor Director:

Al parecer, el doctor Dickerson no ha captado plenamente las profundas implicaciones de la teoría de la "panspermia dirigida" propuesta por Francis Crick y Leslie Orgel, según la cual la vida terrestre pudo haber sido sembrada deliberadamente por seres extraterrestres dotados de inteligencia. A pesar de lo que Dickerson pueda creer, esta teoría sí está relacionada con el examen que se realiza en su artículo, ya que invalida su afirmación de que "si este tipo de vida (la terrestre) no sufrió su evolución inicial en la tierra, seguramente lo hizo en un planeta no muy diverso del nuestro en cuanto a su temperatura y composición". En la escena que nos presenta la teoría de la panspermia dirigida, los seres extraterrestres pudieron muy bien haber sido una forma de vida totalmente diferente de la nuestra (por ejemplo, sin ADN, ni ARN, ni tampoco proteínas), surgida en condiciones que nunca se dieron en la tierra. Estos extraterrestres inteligentes, al llegar a tener conocimiento de la temperatura y composición de nuestro planeta, habrían conseguido, en sus sofisticados laboratorios de investigación en exobiología, diseñar y sintetizar un procariota artificial (con ADN, ARN, proteína y todo lo demás) que tuviera grandes probabilidades de prosperar en el medio ambiente terrestre. Habrían lanzado este artefacto de laboratorio desde su planeta hacia el sistema solar, y el resto ya entra en nuestra historia.

En otras palabras, lo que han demostrado Crick y Orgel es que, sin necesidad de apelar a la creación divina, al menos no resulta imposible, desde un punto de vista lógico, suponer que lo que Dickerson intenta explicar en su artículo, es decir, la evolución química de la única forma de vida que *nosotros* conocemos, de hecho no se hubiera dado nunca.

GUNTHER S. STENT

Universidad de California
Berkeley

Señor Director:

La primera parte de la carta del profesor Fox es una información más que una pregunta, por lo que no precisa respuesta. Me doy por rectificado en lo que concierne a la posibilidad, ya publicada, de formar protocélulas a temperatura ambiente partiendo de un material de proteinoides termales. Parece que entre los científicos que se ocupan del origen de la vida existe cierta controversia en torno al tema de si los proteinoides termales presentan o no de forma significativa cierto grado de autoordenación y un comportamiento no aleatorio; véanse *Molecular Evolution and the Origin of Life*, citado por el profesor Fox, y *The Origins of Life on the Earth*, de S. L. Miller y L. E. Orgel.

Quiero aprovechar esta ocasión para precisar que la bibliografía que acompañaba a mi artículo, y que fue reducida por los editores de *Investigación y Ciencia* al hacer imprimir el artículo, incluía los dos libros anteriormente mencionados como también *The Origin of Life*, de J. D. Bernal (World Publishing Company, 1967) y *Genesis and Evolutionary Development of Life*, de A. I. Oparin (Academic Press, 1969).

El doctor Stent está absolutamente en lo cierto cuando afirma que no podemos rechazar la teoría de que la tierra fue "sembrada" por seres extraterrestres que se sirvieran de principios estructurales, metabólicos y genéticos bastante diferentes de los suyos propios. Del mismo modo, y para citar una paradoja lógica todavía más antigua, tampoco podremos rechazar nunca la teoría de que nuestro universo fue creado hace cinco milisegundos con todas las cosas colocadas en su sitio, incluidas unas "pseudomemorias" bastante espurias en nuestras mentes, que nos hacen pensar que el universo tuvo una larga existencia previa. Aunque estas teorías no son lógicamente rechazables, no obstante, la información de que disponemos actualmente no parece exigir la sugerencia de ninguna de estas hipótesis. Se pueden desarrollar mentalmente unas condiciones y un camino para la evolución independiente de la vida sobre este planeta construyendo así un montaje escénico plausible en líneas generales y, por otra parte, creer (¿quizás entre aquí la fe?) que las lagunas actuales en el conocimiento de este montaje son tales que con toda probabilidad podrán ser colmadas por la investigación sistemática. Claro que tal vez estemos equivocados. Unas pruebas claras de la participación extraterrestre o de la intervención divina elevarían el problema a un nivel de interés intelectual todavía superior.

Cuando hablamos de acontecimientos que se dieron hace tres mil millones de años, nunca podemos decir: "Esto fue lo que ocurrió...", sino tan sólo "Esto probablemente fue lo que ocurrió, por las siguientes razones...". Dentro del ámbito en que entendemos actualmente al problema del origen de la vida, no parecen ser obligadas las huellas extraterrestres en la arena de la historia.

Y para acabar, Thomas Jukes me ha escrito para recordarme que, aunque el molibdeno pueda ser escaso en la corteza terrestre tomada en su conjunto, se halla lejos de escasear en el agua de mar, en la que se encuentra tan comúnmente como el cinc o el hierro. Por tanto, si suponemos que la vida halló su evolución inicial en los océanos, no era tal la escasez de molibdeno que preocupaba a Crick y a Orgel en su publicación sobre la panspermia dirigida.

RICHARD E. DICKERSON

Instituto de Tecnología de California
Pasadena.

Señor Director:

Me he dado cuenta de que en mi trabajo "Poliformismo cromosómico y evolución", que aparece en el número extraordinario, del mes de noviembre, de *Investigación y Ciencia* hay un error en la fórmula de la página 100.

La fórmula dice:

$$D = \frac{1}{2r} \sum_{j=1}^{r} \sum_{k=1}^{sj} (P_{1jk} - P_{2jk})$$

y debería ser:

$$D = \frac{1}{2r} \sum_{j=1}^{r} \sum_{k=1}^{sj} | P_{1jk} - P_{2jk} |$$

o sea, que $P_{1jk} - P_{2jk}$, que aparece entre paréntesis, debe estar entre dos barras, ya que corresponde al valor absoluto de la diferencia, pues al estar entre paréntesis su valor es siempre 0.

ANTONIO PREVOSTI

Catedrático de Genética
Facultad de Biología
Barcelona

Señor Director:

En el número de julio, y dentro de la sección "Juegos matemáticos" de Martin Gardner –traducción de Luis Bou–, se da el desarrollo correcto de:

$$e^{e^x} + (1 = \frac{x}{1!} + \frac{2x^2}{2!} + \frac{5x^3}{3!} + \ldots),$$

y posteriormente, la fórmula de Dobinski, para el n-ésimo número de Bell, B_n,

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}$$

Es de observar que la citada fórmula de Dobinski, y aún suponiendo que k^n se refiera a la n-ésima derivada de la fundación e^{e^x}, la inclusión del signo Σ sumatorio y el factorial de k, no nos conducen a la obtención de los números de Bell; estimo pues que la fórmula correcta sería:

$$B_n = \frac{1}{e} f^{(n)}(x)_{x} = 0$$

siendo $f^{(n)}(x)$ la derivada n-ésima de la función e^{e^x} en el entorno de $x = 0$.

FRANCISCO JAVIER CAMARA
Doctorando de la Facultad de Ciencias
Universidad de Valladolid

Los autores

EDWARD D. GRIFFITH y ALAN W. CLARKE ("La producción mundial de carbón") son economistas especializados en consumo y demanda de energía. Griffith es consultor jefe para análisis de política y prospectiva en la División de Planificación Corporativa de la Atlantic Richfield Company. Estudió economía en la Universidad de Cornell y en la Escuela Londinense de Economía, y recibió su grado de Master of Business Administration, en finanzas, por la Universidad de Columbia. En la Atlantic Richfield analiza las políticas energéticas en curso y propuestas por el Gobierno, y sus repercusiones sobre la economía y la industria energética. Clarke es el jefe de información y análisis sobre energía de la Shell International Petroleum Company, en Londres. Desde su entrada en la Shell, en 1952, ha ocupado diversos puestos relacionados con el análisis y la prospectiva de la demanda. Estadístico por formación, está desarrollando actualmente una "base de datos mundiales sobre consumo energético", con énfasis sobre las regiones en vías de desarrollo.

HARVEY F. LODISH y JAMES E. ROTHMAN ("La formación de membranas celulares") han colaborado en investigaciones sobre la membrana celular. Lodish enseña biología en el Instituto de Tecnología de Massachusetts y es asesor científico en hematología y oncología en el Children's Hospital Medical Center, de Boston. Se graduó en el Kenyon College en química y matemáticas y obtuvo su doctorado en la Universidad Rockefeller en 1966. Después de dos años como becario postdoctoral en el Medical Reseach Council Laboratory of Molecular Biology, de Cambridge, se incorporó al MIT. Rothman es profesor asistente de bioquímica en la Universidad de Stanford. Estudió física en el Yale College y en su año de graduación se dedicó por entero a la investigación biofísica. Rothman permaneció dos años en el laboratorio de Lodish como becario postdoctoral antes de trasladarse a la Universidad de Stanford.

AMNON YARIV ("La óptica de ondas guiadas") es profesor de ingeniería y de física aplicada en el Instituto de Tecnología de California. Nació en Tel Aviv en 1930 y es un veterano de la guerra de la independencia de Israel. En 1951 se trasladó a los Estados Unidos para estudiar ingeniería en la Universidad de California en Berkeley, obteniendo el grado de doctor en 1958. Al año siguiente, se incorporó al equipo de investigación de los Laboratorios Bell y trabajó en el desarrollo de varios de los primeros sistemas de láser de estado sólido. Las actuales investigaciones de Yariv se llevan a cabo en los campos de la óptica no lineal, láseres de electrones libres, láseres de semiconductores y óptica integrada.

THOMAS H. JORDAN ("La estructura profunda de los continentes") profesa geofísica en la Scripps Institution of Oceanography de la Universidad de California en San Diego. En 1972 se doctoró en geofísica y matemáticas aplicadas en el Cal Tech. Su interés investigador se centra en los campos de la sismología y de la tectónica; en particular, se propone "desentrañar los procesos dinámicos que tienen lugar en el interior de la tierra mediante el estudio sismológico de la estructura del globo".

A. TERRY BAHILL y LAWRENCE STARK ("Las trayectorias de los movimientos bruscos del ojo") son bioingenieros que están particularmente interesados en el control del movimiento. Bahill es profesor adjunto de ingeniería eléctrica y bioingeniería en la Universidad Carnegie-Mallon y de neurología en la Facultad de Medicina de la Universidad de Pittsburgh. En calidad de director del laboratorio de Sistemas de Control Neurológico de Carnegie-Mellon, Bahill se dedica al estudio del control de los movimientos oculares y de la locomoción en el hombre. También está investigando sobre el desarrollo de la marcha en los niños, utilizando a su hijo de un año para llevar a cabo el estudio piloto. Stark enseña óptica fisiológica e ingeniería en la Universidad de California en Berkeley, así como neurología en la Universidad de California en San Francisco. Inició sus estudios en la Universidad de Columbia y recibió el título de doctor en medicina en 1948 por el Albany Medical College. Las investigaciones de Stark se centran en el análisis por computador del control del movimiento del ojo y de la mano.

GABRIEL WEINREICH ("Vibraciones acopladas en las cuerdas del piano") es profesor de física en la Universidad de Michigan. Estudió en la de Columbia, realizando su doctorado en física con I.I. Rabi. Después de permanecer siete años en los Bell Laboratories se incorporó a la Universidad de Michigan en 1960. Sus investigaciones han cubierto muchos temas, incluyendo la interacción de las ondas acústicas con los electrones en los semiconductores, lo que le condujo al desarrollo de la amplificación acústica.

MALCOLM R. CLARKE ("La cabeza del cachalote") es un científico del laboratorio de la Marine Biological Association of the United Kingdom en Plymouth. Estudió zoología en la Universidad de Hull, y cuando se graduó empezó a estudiar el tubo digestivo del cachalote. Su tesis doctoral le supuso trabajar como inspector ballenero del gobierno durante ocho meses a bordo de un barco factoría ballenero en el Antártico. De esta experiencia escribe: "Mis primeros intentos de estudiar la histología del intestino del cachalote resultaron frustrados por el enorme tamaño del órgano: Un cachalote de 15 metros tiene un tubo digestivo de más de 200 metros de longitud. Me dediqué a obtener parásitos de estos cetáceos, incluyendo solitarias del conducto biliar de casi 20 metros de longitud".

DAVID LURIÉ y JORGE WAGENSBERG ("Termodinámica de la evolución biológica") enseñan física en los departamentos de Física Teórica y Termología de la Universidad de Barcelona. Lurié cursó sus estudios en la Universidad de Bruselas, donde obtuvo el grado de doctor en 1960. Pasó cuatro años de postdoctorado en el CERN y en las universidades de Chicago y Rochester; posteriormente ocupó plazas docentes en el Dublin Institute for Advanced Studies y en el Technion de Haifa. Proviene del campo de las partículas elementales –es autor de *Particles and Fields*– y su dominio de investigación actual es la termodinámica y mecánica estadística del no-equilibrio. Wagensberg nació en Barcelona en cuya Universidad se doctoró en 1976 con una tesis sobre los aspectos termodinámicos de la evolución de cultivos bacterianos. Recientemente ha formado un grupo de investigación en biofísica dedicado al estudio de los procesos biológicos bajo la óptica de la termodinámica de los procesos irreversibles en colaboración con diversos centros del Consejo Superior de Investigaciones Científicas.

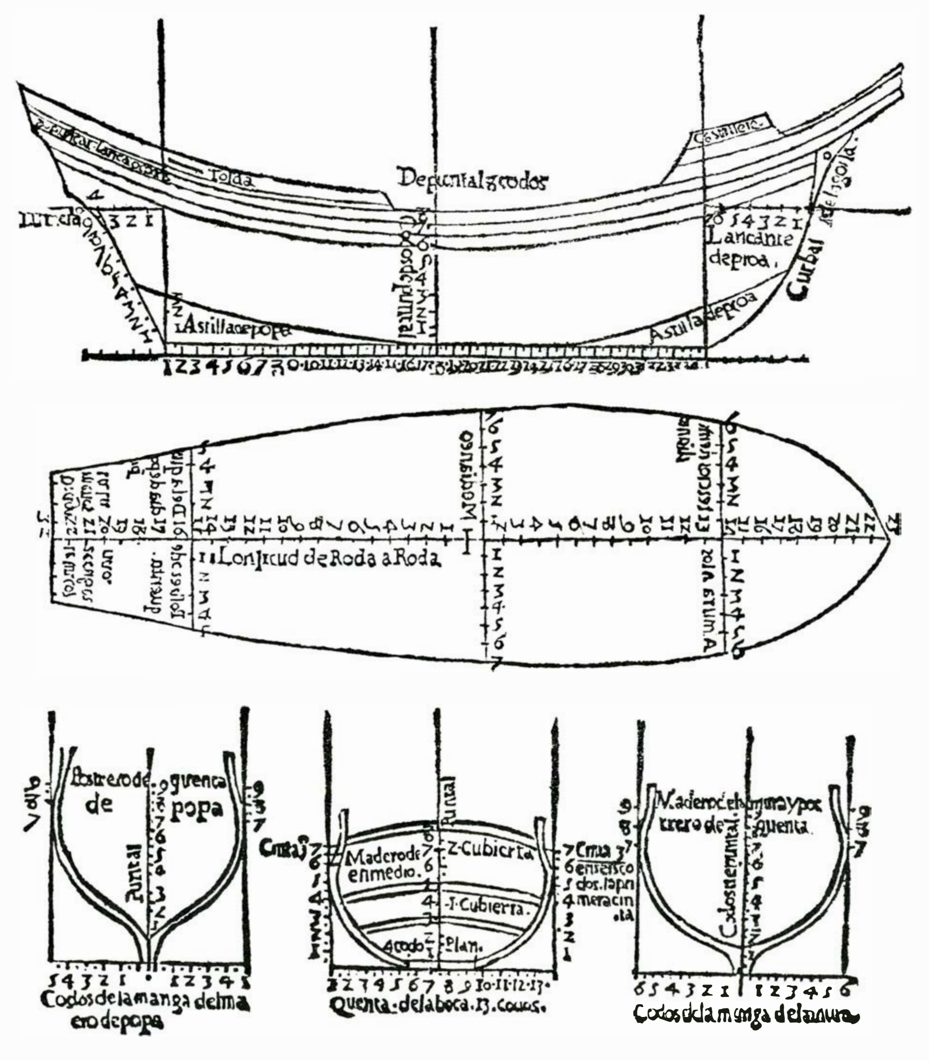
Hace...

José M.ª López Piñero

...cuatrocientos años

El santanderino Diego García de Palacio fue nombrado oidor de la Audiencia de Guatemala. No destacó como un funcionario escrupuloso, ya que llegó a ser procesado a causa de sus abusos. Por el contrario, fue el más importante de los tratadistas sobre temas tecnológicos en la América del siglo XVI. Sus principales obras son las tituladas *Diálogos militares* (1583) e *Instrucción nauthica* (1587), ambas publicadas en la ciudad de México por el impresor Pedro Ocharte. En los *Diálogos* se exponen, por un lado, los conocimientos científicos que entonces se consideraban necesarios para la profesión militar, en especial matemáticas y astronomía. Por otro, una amplia serie de cuestiones técnicas relacionadas con el manejo y la fabricación de las armas y la pólvora. El más célebre de sus capítulos es el dedicado al diseño y fundición de piezas de artillería.

La *Instrucción nauthica* es un estimable texto sobre la materia, aunque sin llegar a la altura de los grandes tratados de Pedro Medina (1545) y Martín Cortés (1551), que habían impuesto en toda Europa el nuevo "arte de navegar". Incluye, como era habitual, un resumen de astronomía teórica e instrucciones para el manejo de los instrumentos de observación y de las tablas, estas últimas calculadas para la latitud de la ciudad de México. Particular interés tiene su cuarto "libro", consagrado a la "traça de naos" y a su maniobra, arboladura y velamen, que fue el primer texto impreso en el mundo sobre construcción naval. Cuida de dar "claras reglas para su proporción", ofreciendo proyectos en los que se detallan minuciosamente las medidas, porque "de abrir o cerrar más de lo necesario cualquier nave, o de no tener el lanzante, delgados y maderos de cuenta, rasales y escora en su punto, salen de mal gobierno y con otras malas condiciones de que suelen venir muchas pérdidas, daños y muertes". La obra lleva como apéndice un "Vocabulario de los nombres que usa la gente de mar", que es asimismo el más antiguo glosario náutico impreso.



Ejemplos de las figuras con medidas para la construcción de barcos que aparecen en la obra de García de Palacio **Instrucción nauthica, para el buen uso, y regimiento de las Naos, su traça, y govierno**

...trescientos años

Falleció el valenciano José de Zaragoza y Vilanova, uno de los más destacados *novatores* responsables de la introducción en España de la Revolución Científica del siglo XVII, en el terreno de los saberes matemáticos y astronómicos. Las obras matemáticas de Zaragoza –entre las que destaca una *Trigonometría española* (1663) que incluye las primeras tablas de logaritmos publicadas en España– respondieron a una intención didáctica. Aunque contribuyeron notablemente a superar el bajo nivel local de la disciplina, no hay que olvidar que tenían algunos graves defectos de información, sobre todo relativos a los progresos del álgebra.

Mucho más valiosa y mejor conectada con las nuevas corrientes europeas estuvo la labor astronómica de Zaragoza. En primer lugar, fue un sobresaliente astrónomo práctico, que diseñó y construyó numerosos instrumentos, entre ellos varios anteojos. Sus observaciones culminaron en las de los cometas de 1664 y 1677. El informe de la correspondiente al primero fue traducido al francés al cabo de más de un siglo (1783). Su observación del cometa de 1677 fue la primera de Europa, siendo recogida en el *Journal des Savants* y en las memorias de la Academia de Ciencias de París. En segundo término, Zaragoza tuvo una información rigurosa y al día de la ciencia astronómica de su tiempo. Ello le permitió publicar su libro *Esphera en común, celeste y terráquea* (1675), el mejor tratado astronómico aparecido en la España del siglo XVII. Su orientación es claramente moderna, ante todo por la continua fundamentación de las hipótesis en datos de observación, en buena parte propios. También lo es por la explícita aplicación de los nuevos datos a la superación de las doctrinas de la cosmología clásica. La única limitación de importancia es la cuestión del heliocentrismo y del movimiento de la Tierra, sobre la que pesaba una prohibi-

ción mantenida con especial energía por la Inquisición. El detenido análisis interno que Cotarelo dedicó a la obra de Zaragoza le hizo sospechar, sin embargo, que el astrónomo valenciano fue "copernicano en secreto".

...doscientos años

José Ignacio Bartolache publicó su *Instrucción que puede servir para que se cure a los enfermos de las viruelas*, con motivo de una epidemia que entonces asolaba a Nueva España. Nacido en Guanajuato y formado como médico en la Universidad de México, Bartolache contribuyó de manera destacada a la renovación de la medicina novohispana durante la segunda mitad del siglo XVIII. Fue seguidor de la mentalidad antisistemática que, partiendo de la obra del holandés Boerhaave, desarrolló principalmente la primera escuela de Viena. Por ello, procuró incorporar cuantas novedades le parecieron útiles, al margen de cualquier prejuicio doctrinal. Defendió, por ejemplo, la inoculación de las viruelas e introdujo el empleo del nuevo medicamento llamado "fierro sutil". Preocupado por la difusión de este último entre los indios, tradujo al náhuatl la *Instrucción* que le había dedicado. Bartolache publicó asimismo la primera revista médica que apareció en América. Con el título de *Mercurio volante, con noticias importantes y curiosas sobre varios asuntos de física y medicina*, alcanzó dieciséis números (1772-1773), en los que figuraron, entre otros, trabajos dedicados al barómetro y al termómetro, a la histeria y a la bebida mexicana llamada "pulque".

A pesar de todo, Bartolache nunca cultivó la medicina con entusiasmo. Se sintió especialmente atraído por las ciencias físicas y matemáticas, en las que realizó también alguna aportación de interés. Anotaremos únicamente el primer cuaderno de sus *Lecciones matemáticas* (1769) y la observación del tránsito de Venus por el disco del Sol, en colaboración con José Antonio Alzate.

...cien años

Apareció en Madrid el trabajo de Arturo Soria y Mata *Avisador de las crecidas de los ríos*, en el que exponía un nuevo sistema de indicadores enlazados eléctricamente cuya finalidad era prevenir las inundaciones. Su autor fue uno de los ingenieros más originales de la España de finales del siglo XIX y comienzos del actual. Buena parte de sus invenciones estuvo relacionada, de una forma u otra, con el problema de la comunicación. Diseñó un "teodolito-impresor automático" (1870); instaló el primer tranvía madrileño, con la idea de que la principal función de los transportes urbanos era enlazar las estaciones con los mercados; propuso también el "ferrocarril-tranvía de circunvalación".

Arturo Soria fue todo lo contrario de un ingeniero de mentalidad científica. El apoyo doctrinal de su obra fue, por una parte, una visión geométrica de la realidad inspirada en la cristalografía de Cauchy. A diferencia de este autor, Soria pensaba que el tetraedro era el único poliedro regular, del que derivaban todos los demás. Por otra parte, era un convencido partidario del evolucionismo de Darwin y Spencer, que intentó desarrollar, entendiendo como sistemas geométricos, tanto la realidad inorgánica como los seres vivos. Estos fueron los supuestos de su formulación de la "ciudad lineal", la más famosa de sus aportaciones. Como sistema de comunicación "espinal", la consideró más evolucionada que las agrupaciones urbanas tradicionales, que serían "invertebradas", siguiendo con la analogía biologista.

...cincuenta años

Más de un centenar de biólogos de todo el mundo, entre ellos los entomólogos de mayor prestigio internacional, ofrecieron a Ignacio Bolívar Urrutia un volumen en homenaje a su labor científica, con motivo de su cercana jubilación. Cuando contaba poco más de veinte años, Bolívar había sido uno de los fundadores de la Sociedad Española de Historia Natural (1871). Obtuvo la plaza de ayudante del Museo de Ciencias Naturales en 1875 y la de catedrático de entomología de la Universidad de Madrid en 1877. Su obra personal de investigador se centró en la entomología, especialmente en los ortópteros y los hemípteros, sobre los que publicó trabajos en las mejores revistas científicas europeas, describiendo por vez primera numerosos géneros. También ordenó y analizó los materiales recogidos en diversas expediciones científicas, entre los que destacan los procedentes del viaje al Pacífico desde 1862 a 1865, que se conservaban en el Museo de Historia Natural de Madrid. Por su autoridad en la materia, fue encargado de revisar regularmente las colecciones ortopterológicas de los principales museos europeos.

Otra contribución de Bolívar, de tanto relieve como su labor individual, es que supo convertir el Museo de Ciencias Naturales y su laboratorio en la Universidad de Madrid en uno de los núcleos más exigentes y al día de la investigación biológica en España.

La producción mundial de carbón

Cuando la producción mundial de petróleo alcance su techo y comience a decrecer, se necesitarán enormes cantidades de otras fuentes sustitutivas. El carbón puede cubrir una parte importante de las necesidades energéticas

Edward D. Griffith y Alan W. Clarke

Antes de la Segunda Guerra Mundial, el carbón era el combustible predominante en todo el mundo. Desde entonces ha cedido terreno a otros combustibles, y actualmente sólo suministra el 19 por ciento de la energía consumida por los países no comunistas. Durante los últimos 20 años, mientras el consumo de petróleo se triplicaba en todas partes, la producción de carbón permanecía constante, a excepción de en la Unión Soviética, China y Polonia. En estos países, la producción de carbón aumentó un 30 por ciento entre 1960 y 1977; sin embargo, también quedó retrasada con respecto al consumo energético total. El virtual estancamiento del consumo de carbón en los países no comunistas refleja la preferencia de los consumidores, industriales y privados, por la comodidad que ofrecen los combustibles líquidos, el gas natural y la electricidad, y también el crecimiento de los mercados específicos del petróleo, como son los del transporte y los petroquímicos. Con el reconocimiento generalizado de que la producción de petróleo está probablemente estabilizándose, y de que por otro lado decrecerá durante los próximos 25 años, las naciones se verán obligadas a buscar combustibles alternativos. Una vez más, en todo el mundo, se echará mano del carbón para mantener la producción industrial y el desarrollo económico.

Para conseguir incrementos notables en la producción y en la utilización del carbón, se precisarán fuertes incentivos y cambios fundamentales de actitudes, ya que el carbón no es tan fácil de extraer, transportar y quemar como lo son el petróleo y el gas natural. La disponibilidad y la aceptación del carbón se lograrán mediante una importante ampliación de los medios de explotación, tratamiento, transporte y utilización del mismo, y también mediante unas nuevas tecnologías de tratamiento, conversión y consumo de carbón. De acuerdo con los proyectos más modestos para el año 2000, Estados Unidos tendrá que extraer un 60 por ciento más de carbón que el que extrae hoy. Probablemente habrá de triplicar su extracción. Ante una expansión de tales proporciones, deberán resolverse los serios problemas económicos, sociales y ambientales que están implicados en la minería y la combustión del mismo.

Esbozaremos las consecuencias de un desarrollo del carbón a gran escala, como alternativa de los combustibles fósiles líquidos, las opciones posibles y las vías de solución que habrán de evaluar los políticos para hacer frente al problema en los próximos 25 años. Nuestras estimaciones de demanda y de suministro de carbón se basan en un estudio de dos años de duración, el Seminario sobre Estrategias Energéticas Alternativas, elaborado bajo la dirección de Carroll L. Wilson, del Instituto de Tecnología de Massachusetts. El estudio implicó a unos 70 especialistas de industrias, gobiernos y universidades de 12 países consumidores de energía (Estados Unidos, Canadá, Gran Bretaña, Francia, Alemania Occidental, Holanda, Dinamarca, Suecia, Noruega, Finlandia, Italia y Japón) y de tres importantes países exportadores de petróleo (Irán, México y Venezuela).

El Seminario desarrolló una serie de proyectos energéticos, para los años 1985 y 2000, basados en diversas hipótesis de crecimientos económicos, precios energéticos y políticas gubernamentales. La crítica conclusión extraída de tales proyectos fue que, antes o después, y probablemente antes de que finalice el siglo, la demanda de petróleo pondrá en apuros al suministro mundial, y que la producción global de petróleo alcanzará su techo y luego decrecerá. En estas páginas apareció una detallada descripción de los estudios hechos sobre el petróleo por el Seminario [véase "La producción mundial de petróleo", por Andrew R. Flower: INVESTIGACION Y CIENCIA, mayo, 1978]. En el presente artículo consideraremos la posible utilización del carbón como sustituto del petróleo, la cuantía de recursos mundiales de carbón, las perspectivas de su demanda, su producción potencial en varias regiones del mundo y la posible conversión del mismo en combustibles sintéticos.

Aunque un pedazo de carbón puede parecer exactamente igual que otro, hay muchos rangos y calidades diferentes, como también existen muchos tipos diferentes de crudos petrolíferos; y así un recuento de carbón puede diferir mucho de otro. Los carbones de rangos superiores tienen un poder calorífico de unas 7000 kilocalorías por kilogramo; los de rangos inferiores, la mitad o la tercera parte de dicha cifra. Un sistema de clasificación de los carbones los divide en dos grandes categorías: carbones duros (antracitas y hullas) y carbones blandos (lignitos pardos y negros). La aptitud de los diversos tipos de carbón para formar coque, cuando se calientan en ausencia de aire, es muy variable. Ya que el coque siderúrgico es una de las mayores aplicaciones del carbón, la segunda clasificación importante de los carbones los divide en: siderúrgicos (o coquizables) y térmicos (o de vapor), que se queman para obtener calor o vapor destinado a la generación de energía eléctrica.

La producción mundial de carbón de todos los tipos en 1977 fue de 3400 millones de toneladas métricas, de las que 2500 millones, o sea, un 70 por ciento, eran carbones duros (antracitas y hullas). Los principales productores de carbones duros son los países más desarrollados. Más del 75 por ciento de la producción mundial se extrae por seis países: Estados Unidos, Unión Soviética, China, Polonia, Gran Bretaña y Alemania Occidental. Entre los países en vías de desarrollo, el único productor impor-

tante es la India. Los Estados Unidos contabilizan más de la mitad del carbón extraído y consumido por los países no comunistas. La Unión Soviética y China producen, cada uno, el 80 por ciento de los Estados Unidos.

La producción mundial de lignito pardo y de lignito negro se ha incrementado paulatinamente, alcanzando los 950 millones de toneladas métricas en 1977. Alemania Oriental es el mayor productor, con 250 millones de toneladas, seguido por la Unión Soviética con 180 millones de toneladas. De los países no comunistas, Alemania Occidental es el mayor productor de lignito pardo y lignito negro. Por el menor contenido energético de estos carbones, los lignitos sólo contribuyen con el 4 por ciento del total global de energía primaria en 1977, frente al 25 por ciento de la misma que atendieron los carbones duros.

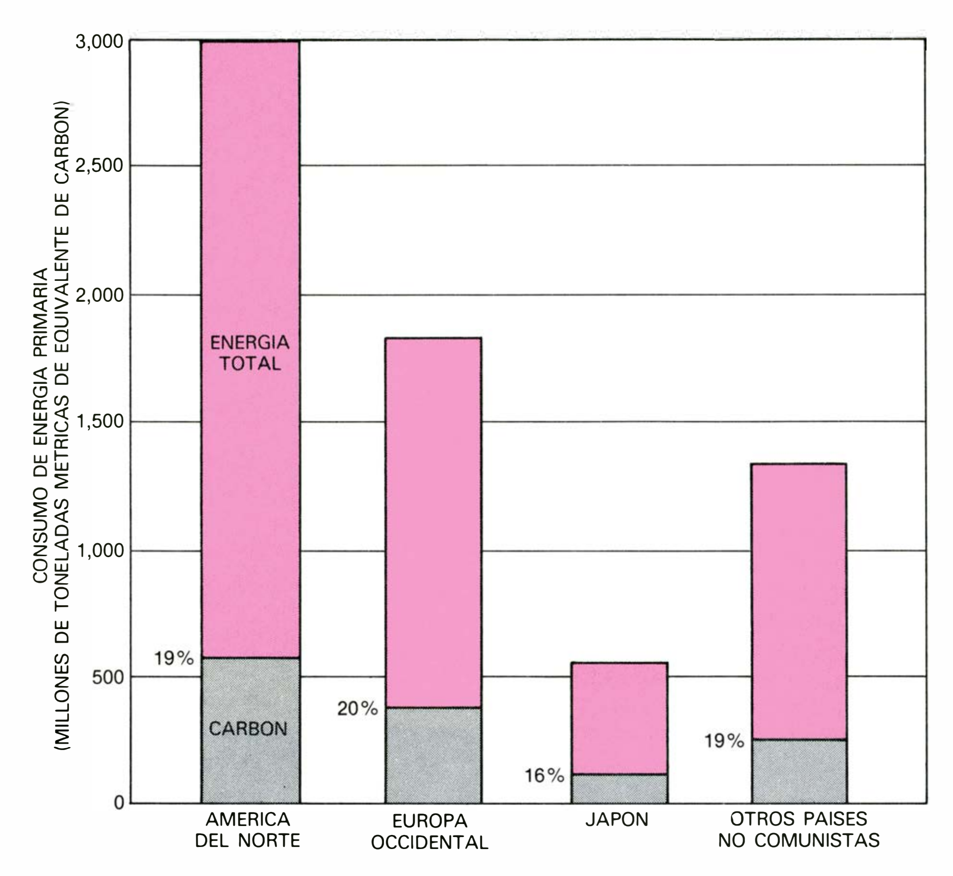
En los países no comunistas, la parte de energía primaria debida al carbón (93 por ciento de ella a carbones duros) disminuyó desde el 33 por ciento en 1960 hasta el 19 por ciento en 1976. La utilización del carbón ha decaído constantemente en todos los mercados, excepto en el de producción de electricidad, que actualmente representa más del 75 por ciento del carbón consumido en los Estados Unidos, el 57 por ciento en Europa Occidental, el 18 por ciento en Japón y el 36 por ciento en todos los demás países del bloque no comunista.

La mayor parte del carbón se consume en los países que lo extraen, al contrario de lo que ocurre con el petróleo, cuyas dos terceras partes se exportan desde los países productores. En 1976, sólo el 8 por ciento de la producción mundial de carbón duro (antracitas y hullas), unos 190 millones de toneladas, se movieron en el mercado internacional, la mayor parte como carbón siderúrgico. El mayor importador fue Japón, que compró 50 millones de toneladas, la mitad de ellas procedentes de Australia.

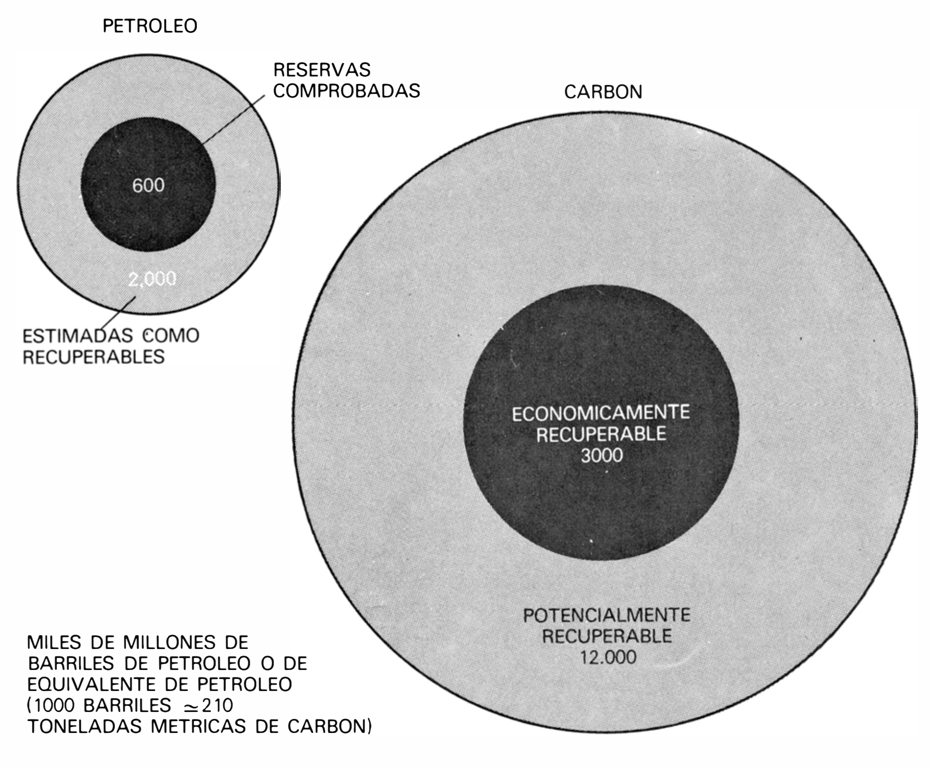
Las reservas mundiales de carbón exceden con mucho a las de cualquier otro combustible fósil, y son suficientes para soportar perfectamente un masivo incremento del consumo durante el próximo siglo. El *Estudio de Recursos Energéticos*, de la Conferencia Mundial de la Energía de 1976, estimó los recursos globales de carbón en 11,5 billones de toneladas métricas. De este total, tan sólo 1,3 billones de toneladas se computan como reservas conocidas, o reservas cubicadas. A su vez, de esta cifra unos 740.000 millones de toneladas métricas se consideraban como económ-



TRENES "UNIFICADOS" de transporte de carbón. Se desarrollaron para mover este combustible con rapidez y a bajo coste, en general entre una mina y un cliente, que casi siempre es una empresa eléctrica. Los trenes van y vienen sin desenganche alguno, funcionando como una cinta transportadora. El tren unificado de 110 vagones, que aparece en la fotografía, es uno de los cuatro trenes que, en conjunto, entregan 2,6 millones de toneladas de carbón por año, a 965 kilómetros desde una explotación a cielo abierto de la Amax Carbon Products, próxima a Gillette, en el estado de Wyoming, hasta una central térmica de la Public Service Company of Colorado, cerca de Pueblo, en el estado de Colorado. Durante 1975, estos trenes unificados acarrearon, en Estados Unidos, 169 millones de toneladas, o sea, la mitad de la totalidad del carbón transportado por ferrocarril.



CONTRIBUCION MEDIA DEL CARBON, en los suministros de energía primaria de los países no comunistas, que fue del 19 por ciento, en 1976, con pequeñas desviaciones para las principales regiones. En 1960, el carbón suministró el 33 por ciento de la demanda total de energía en los países del bloque no comunista, y un 45 por ciento de la demanda total mundial. En 1976, la participación del carbón en los suministros mundiales, incluyendo a los países comunistas, ha descendido casi al 30 por ciento. La mayor parte de la demanda incremental se ha atendido con otros combustibles y energía nuclear.



COMPARACION de las reservas de petróleo y las de carbón. Dicho cotejo indica que el suministro mundial de carbón, potencialmente recuperable, equivale a unos 12 billones de barriles de petróleo, o sea, a unas seis veces las reservas estimadas de petróleo recuperable. Con tecnología y precios de hoy, la cantidad de carbón que resulta económicamente recuperable es bastante menor; equivale a unos 3 billones de barriles de petróleo, cifra que es unas cincos veces las reservas comprobadas de petróleo.

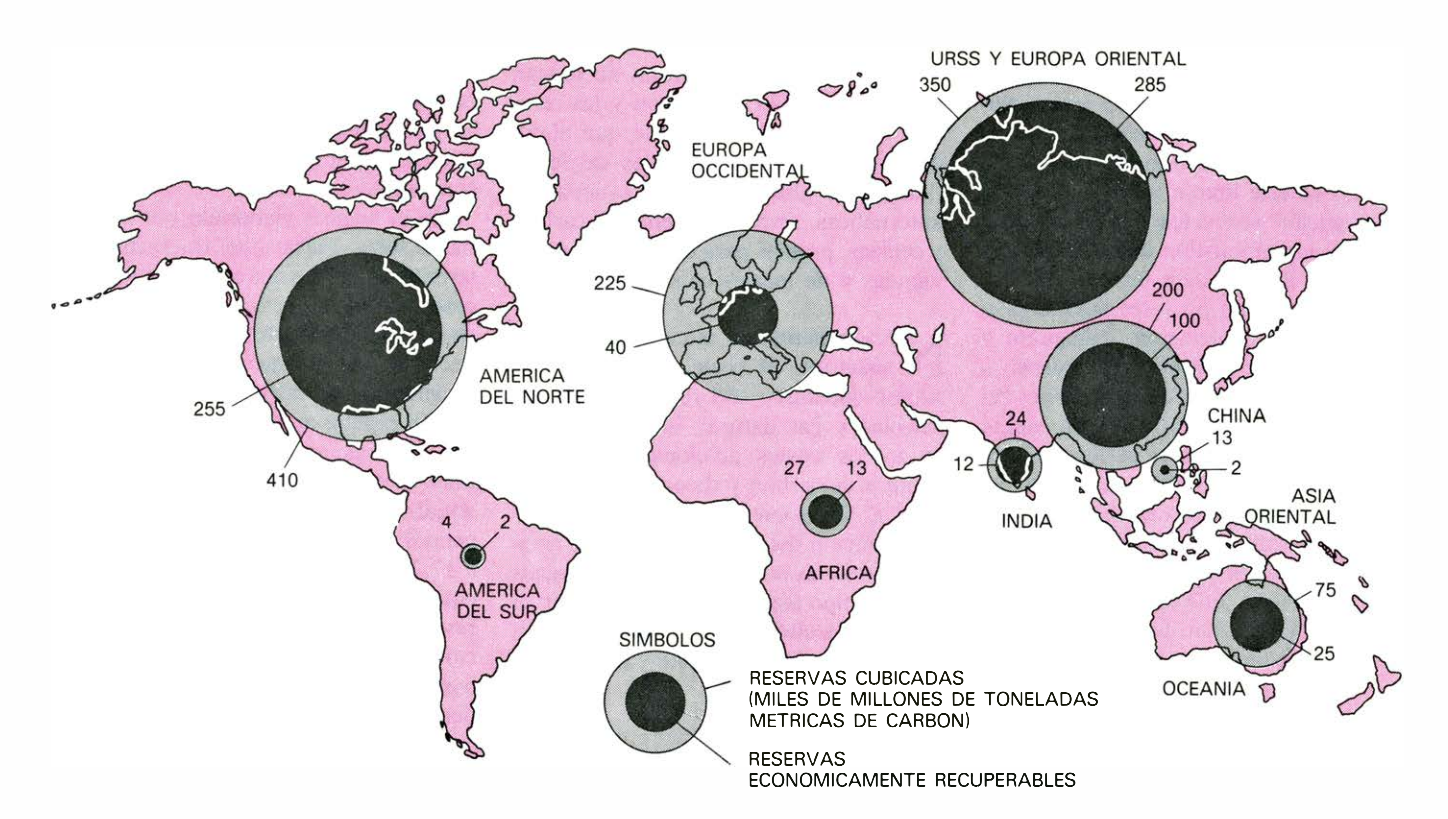
micamente recuperables, cuando se hicieron los estudios del Seminario sobre Estrategias Energéticas Alternativas. El 96 por ciento de las reservas conocidas está en cinco zonas: América del Norte con el 31 por ciento, la Unión Soviética y otros países de Europa Oriental con el 26, Europa Occidental con el 17, China con el 15 y Australia con el 6 por ciento.

La evaluación de la recuperabilidad varía mucho de un yacimiento a otro, pues depende de la accesibilidad del carbón, que está condicionada por muchas variables, tales como el espesor de capa y la profundidad y tipo de terreno. Los porcentajes de recuperación varían del 85 al 95 en explotaciones a cielo abierto, y del 25 al 70 en explotaciones subterráneas.

La estimación hecha por el Seminario, de 740.000 millones de toneladas de carbón económicamente recuperable, incluye todos los tipos de carbón. Cuando se introducen ajustes, por menor poder calorífico de los de rango inferior, el tonelaje recuperable pasa a unos 600.000 millones de toneladas métricas de carbón duro equivalente, suficientes para el consumo de 200 años, a las tasas actuales. Expresadas en términos de petróleo equivalente, los 600.000 millones de toneladas métricas de carbón duro tienen un contenido energético de unos 3 billones de barriles, lo que significa de 4 a 5 veces las estimaciones corrientes de las reservas comprobadas de petróleo.

La estimación de la Conferencia Mundial de la Energía de 11,5 billones de toneladas métricas, como reserva mundial total de carbón, equivale a unos 50 billones de barriles de petróleo. Con un criterio conservador, de recuperar sólo el 25 por ciento de los recursos de carbón, podría recuperarse finalmente un total de 2,5 billones de toneladas de carbón, equivalentes a 12 billones de barriles de petróleo. Esta cifra representa seis veces la cantidad estimada de petróleo recuperable.

Las valoraciones de los recursos mundiales de carbón no son tan completas como las del petróleo, porque sus prospecciones de yacimientos han sido más reducidas y menos intensas. Muchas estimaciones de las reservas de carbón se hicieron cuando el petróleo era abundante y barato, de modo que había poco incentivo para buscar más carbón. Conforme crezca el interés por éste, deberán identificarse más reservas. Por ejemplo, pocos países en vías de desarrollo han realizado prospecciones de carbón. Sin embargo, la geología sugiere que importantes áreas del mundo, inclu-



TOTAL DE LAS RESERVAS cubicadas de carbón; alcanza la cifra de 1,3 billones de toneladas métricas, de las que 740.000 millones se estiman como económicamente recuperables; lo que representa unos 200 años de suministro a las tasas actuales de producción. (Ajustadas por poder calorífico, los 740.000 millones de toneladas equivalen a 3 billones de barriles de petróleo, cifra señalada en la ilustración inferior de la página precedente.) Las "reservas cubicadas" son las cantidades de carbón de los yacimientos perfectamente reconocidos, cuya extensión y calidad se han determinado mediante adecuadas muestras. La mayoría de las reservas de carbón, actualmente cubicadas, se hallan en el hemisferio norte, aunque es de esperar que el resto del mundo incremente sus reservas cubicadas gracias a las nuevas investigaciones y prospecciones de yacimientos de carbón en el hemisferio sur.

yendo gran parte del hemisferio sur, tienen un elevado potencial en carbón.

Australia es un ejemplo de desarrollo reciente en carbón. Durante los últimos 20 años, los geólogos realizaron intensas prospecciones, principalmente inspiradas por las necesidades japonesas de coque siderúrgico. Como resultado, se encontraron importantes reservas de carbón en Nueva Gales del Sur y en Queensland, pasando Australia a triplicar su producción anual. Las reservas australianas de carbón bituminoso pueden compararse, por su contenido energético, a las reservas comprobadas de petróleo de Arabia Saudita. Por prospecciones análogas, en Sudáfrica se están obteniendo resultados favorables, lo que indica que las reservas mundiales de carbón pueden ser mucho mayores que las que han sido estimadas. Por otra parte, hay casi con certeza grandes yacimientos submarinos de carbón; por ejemplo, se piensa en yacimientos importantes bajo el Mar del Norte. Estas potenciales reservas de carbón no se incluyen ni en las estimaciones más optimistas de tales reservas.

Sin embargo, la exacta cubicación de las reservas mundiales de carbón no constituye ningún punto crítico. Los recursos conocidos de carbón pueden soportar cualquier ritmo de explotación probable, durante las próximas décadas. Lo que ya es dudoso es la disposición y la capacidad de las naciones para aceptar importantes aumentos de la producción y la utilización del carbón.

Para determinar su papel futuro, la clave está en calcular el desarrollo potencial de la demanda de carbón. Si la demanda de éste aumenta y se prevé con bastante antelación, para realizar las inversiones requeridas en producción y transporte, podrían obtenerse notables incrementos de producción de carbón. La cuestión crítica está, pues, en si la demanda de carbón continuará estancada o si aumentará precisamente al decrecer la disponibilidad de petróleo.

Antes de que el carbón pueda cubrir de nuevo participaciones crecientes, en las necesidades mundiales de energía, habrán de cumplimentarse una serie de condiciones. Primero, debe tenerse una clara conciencia de que las alternativas del petróleo son necesarias, y de que el carbón constituye uno de los combustibles más baratos y abundantes. Esto implica un conocimiento de las perspectivas energéticas a largo plazo, de las decisiones políticas gubernamentales para incrementar la producción y la utilización del carbón, y de las actitudes públicas que harán factibles tales políticas.

También hay que tener en cuenta que, al generalizarse la combustión de carbón sin controles adecuados, se contaminaría el aire, con serias consecuencias para el ambiente y posiblemente para el clima. Para controlar las emisiones de dióxido de azufre y de óxidos de nitrógeno, procedentes de la combustión de carbón, se están desarrollando nuevas tecnologías. Por otro lado, la combustión del carbón produce más dióxido de carbono por unidad de calor desprendido que cuando se quema petróleo o gas natural, dada la alta relación de carbono a hidrógeno que tienen los carbones. El contenido de la atmósfera, en dióxido de carbono, parece que aumenta, y algunos científicos están preocupados con sus efectos sobre el clima. Se desconoce cuál será la contribución de la combustión de combustibles fósiles, en el incremento del dióxido de carbono atmosférico, ni cuáles sus últimas consecuencias; sin embargo, se debe hacer frente a los posibles efectos perjudiciales de un considerable incremento de combustiones de carbón. Se precisan investigaciones y desarrollos adicionales en desulfuración de humos, en combustión con lechos fluidos, en limpiezas quí-

micas del carbón y en otras técnicas para quemar el carbón con una polución mínima. También deben actuar los gobiernos, promoviendo debates sobre normas de aire limpio y estimulando la investigación sobre los efectos a largo plazo de la combustión de combustibles fósiles, en la atmósfera terrestre.

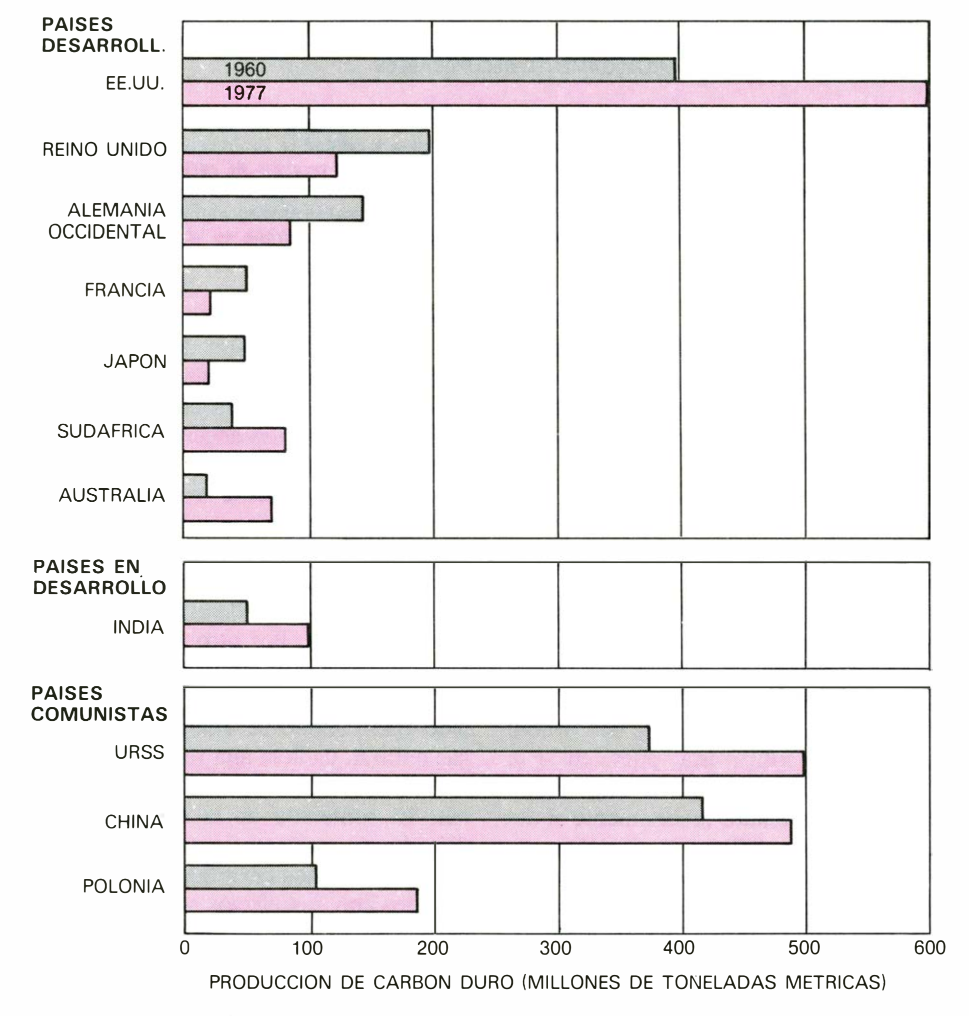
También deben preverse medios progresivos para explotación, transporte y combustión del carbón. La puesta a punto de tales medios e instalaciones requiere cuantiosas inversiones de capital y desarrollo de métodos perfeccionados de manipulación limpia y cómoda del carbón. Unas técnicas mejoradas de manipulación interesan particularmente a los consumidores industriales, comerciales y domésticos, que de otra forma preferirían la utilización de petróleo o gas. Los consumidores de zonas residenciales y de los pequeños comercios no desean volver al paleado del carbón y las cenizas. No obstante, es posible que algún día haya pequeñas unidades de lecho fluido con alimentación y evacuación automáticas, respectivamente de carbón y cenizas, propias para calefacción de viviendas y de locales comerciales.

Normalmente, en el caso de grandes usuarios, el coste del carbón resulta competitivo con relación a los del petróleo y gas natural, incluso computando los costos adicionales para los equipos de carbón y de control de polución. Si los gobiernos desean estimular la utilización del carbón, se pueden establecer unas políticas de precios energéticos y de tipo fiscal que complementen el incentivo económico del empleo del carbón. La elección entre éste y otras fuentes energéticas será consecuencia de una combinación entre precio, condiciones ambientales, comodidad y otros factores que redunden en el propio interés de los consumidores.

En el análisis elaborado por el Seminario sobre Estrategias Energéticas Alternativas, estaba claro que, en todos los casos, podía aumentarse la producción de carbón para cubrir la demanda proyectada. Las estimaciones se refieren a la demanda de carbón por parte de países del bloque no comunista, elaboradas con diversas tasas de crecimiento económico mundial (entre 3 y 6 por ciento anual) en combinación con costes energéticos de diversos precios (desde 11,50 a 17,25 dólares el barril de petróleo o su equivalente energético), con diversas respuestas a las políticas energéticas nacionales ("vigorosa" o "restringida") y con mayor atención por parte de los poderes públicos hacia el carbón o bien hacia la energía nuclear. Los países incluidos consumen actualmente el 70 por ciento de la energía mundial.

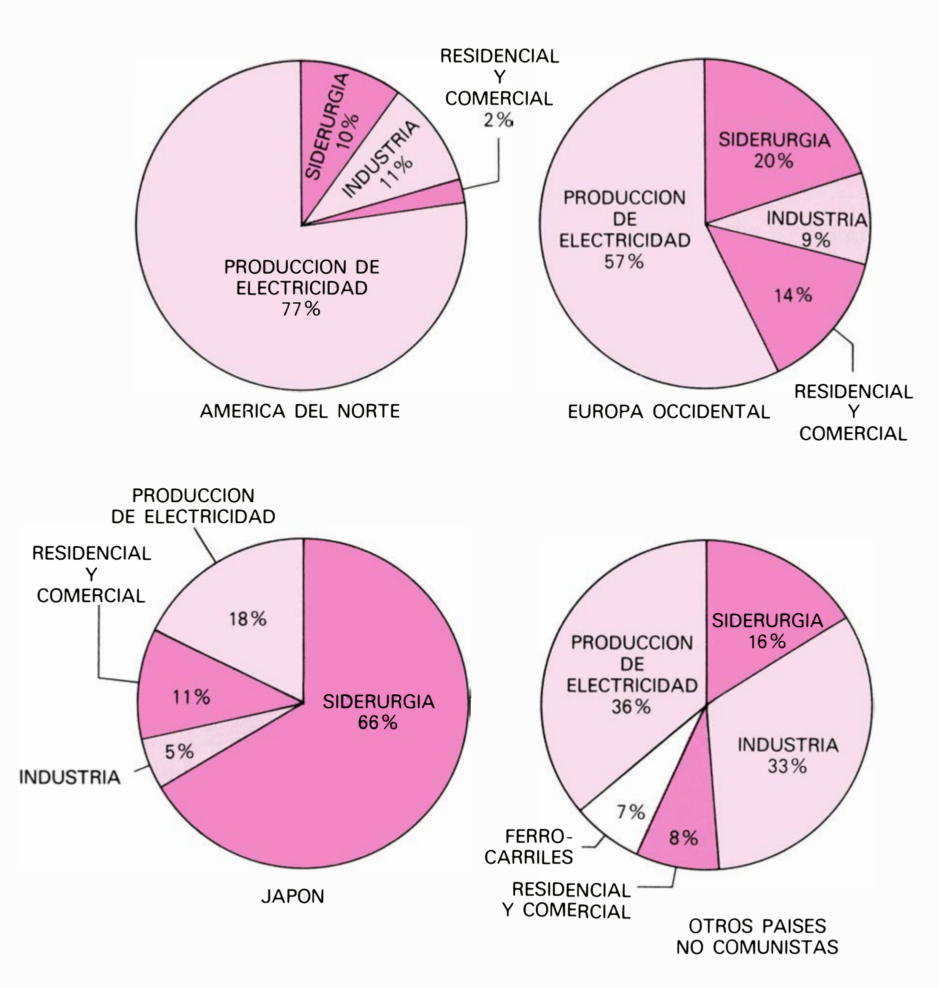
La mínima demanda proyectada de carbón, para el año 2000, es de 1610 millones de toneladas métricas de carbón duro o su equivalente, es decir, sólo un 27 por ciento superior a la de 1976. Este mínimo proyectado se basa en una tasa de crecimiento económico anual del 3 por ciento y en un costo de la energía de 11,50 dólares por barril de petróleo o su equivalente (precio 1975), en combinación con medidas políticas gubernamentales ordenadas a favorecer el desarrollo de la energía nuclear. La mayor demanda proyectada, con 2575 millones de toneladas métricas (un aumento del 103 por ciento sobre 1976), resulta de otro modelo según el cual la tasa de crecimiento anual del 5 por ciento se combina con un costo de la energía de 17,25 dólares el barril de petróleo o su equivalente (un 50 por ciento más que en 1975), y con medidas políticas que ponen más énfasis en el carbón que en la energía nuclear. Con la misma tasa de crecimiento económico y el mismo precio de energía, pero cambiando ahora todo el énfasis en favor de la energía nuclear, la demanda de carbón se eleva todavía a 2225 millones de toneladas, es decir, un 75 por ciento superior a 1976. En este modelo, la demanda de América del Norte sola será de 1035 millones de toneladas, un 85 por ciento de aumento sobre 1976.

Un punto fundamental está en la posibilidad de que el carbón sustituya directamente a los combustibles derivados de los menguantes petróleos, puesto que la producción de carbón puede exceder



PRODUCCION MUNDIAL DE CARBON DURO en 1977. Fue de 2465 millones de toneladas métricas, con un incremento del 3,2 por ciento sobre el año precedente y un 23 por ciento mayor que la cifra del año 1960. La expresión carbón duro excluye los lignitos negros y pardos, combustibles de menor poder calorífico, cuyas producciones alcanzaron unos 910 millones de toneladas métricas en 1977, la mitad de las cuales prácticamente extraídas en Alemania Oriental y la Unión Soviética. La mayoría del incremento de producción de carbón duro, desde 1960, se produjo en los países comunistas. A lo largo de un período de 18 años, la producción total de carbón duro en los países no comunistas fue entre 1100 y 1200 millones de toneladas. Dentro de este total, prácticamente estacionario, hubo una gradual subida de la producción en América del Norte, un descenso de la producción de Europa Occidental y Japón y un aumento notable en las explotaciones de Sudáfrica, Australia y la India.

potencialmente de cualquier demanda proyectada. Nuestros análisis indican que, con un programa de total sustitución del abundante carbón por el limitado petróleo, la demanda de carbón para el año 2000 podría superar las previsiones calculadas en una cifra entre 200 y 600 millones de toneladas. Por ejemplo, en un caso límite, si el carbón se sustituye por una cantidad equivalente de petróleo, la demanda de carbón para el año 2000 sería el doble de la estimada por el Seminario. Aunque resulta incierto el grado de disposición y de capacidad de los usuarios de energía para utilizar carbón en lugar de petróleo, estos análisis subrayan la conveniencia potencial del carbón como sustituto principal del petróleo. Al mismo tiempo, la magnitud del cambio que se requiere nos indica el tamaño de las dificultades a resolver, en el caso de hacer una transición en gran escala del petróleo al carbón.



LOS PRINCIPALES MERCADOS DEL CARBON, en todo el mundo, fueron las centrales termoeléctricas y la industria siderúrgica. Más de las tres cuartas partes del carbón consumido en América del Norte se utilizó para producir electricidad. El principal usuario en Japón fue la industria siderúrgica. Han sido significativas las utilizaciones industriales del carbón en países en vías de desarrollo.

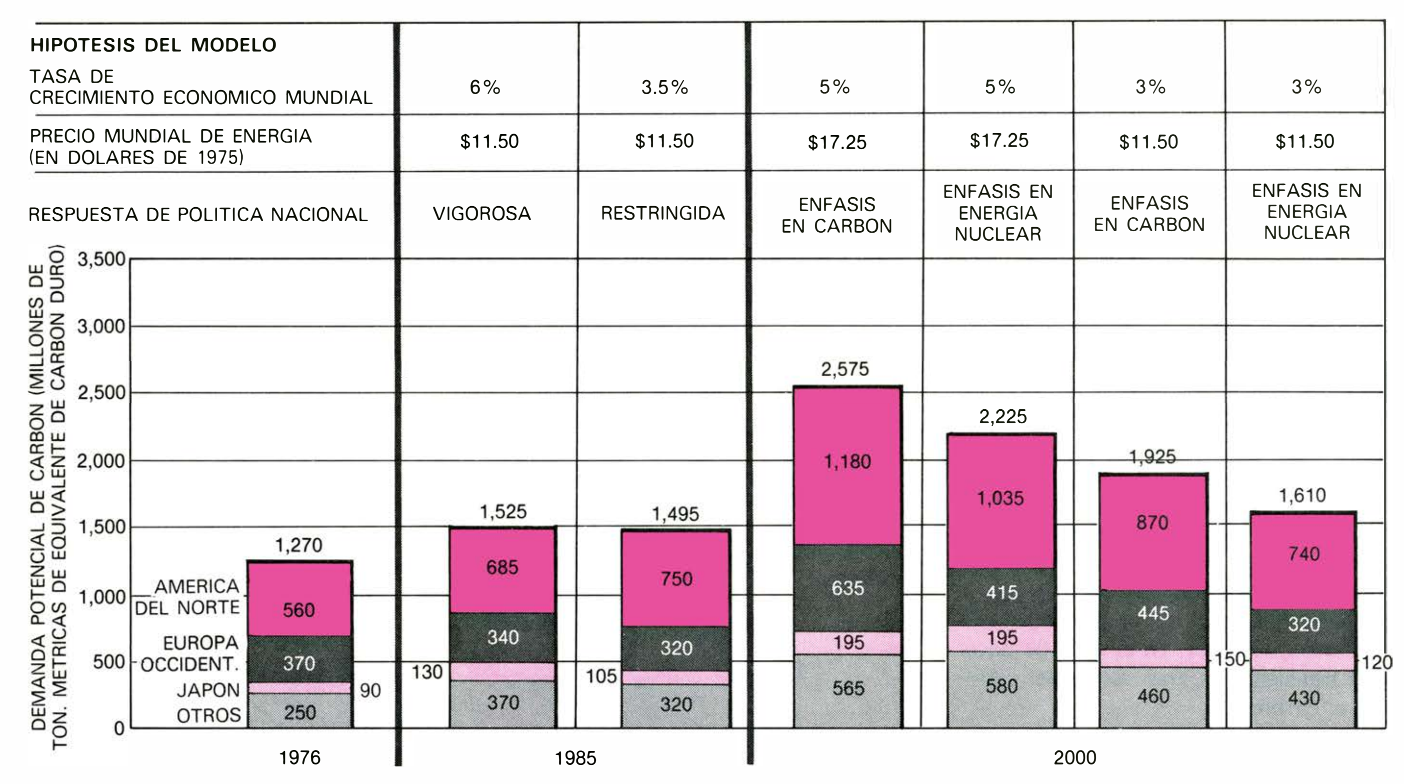
Los distintos gobiernos nacionales tendrán que decidir la extensión en que deben incitar al, o disuadir del, incremento de los sistemas consumidores de carbón. Las medidas precisas deben tomarse pronto, puesto que se necesita mucha anticipación para salir de un sistema energético dominado por el petróleo, si se desea que el cambio se produzca antes de que éste escasee más y se encarezca. Las decisiones tomadas en los años venideros inmediatos marcarán la pauta para el resto del siglo.

En la otra cara de la moneda, opuesta a la de la demanda, está el suministro potencial de carbón, estimado en cifras reales sobre la base de recursos disponibles que permitan la apertura de nuevas explotaciones mineras y la construcción de los medios auxiliares precisos. Cada uno de los grupos nacionales que participó en los estudios del Seminario preparó cuidadosas estimaciones de las producciones potenciales de carbón, para diversos modelos de crecimiento económico, de precio de la energía y de las medidas gubernamentales sobre combustibles. De los países no comunistas que no estaban representados en el grupo de estudio, se hizo también un análisis especial de su producción potencial de carbón y de sus exportaciones.

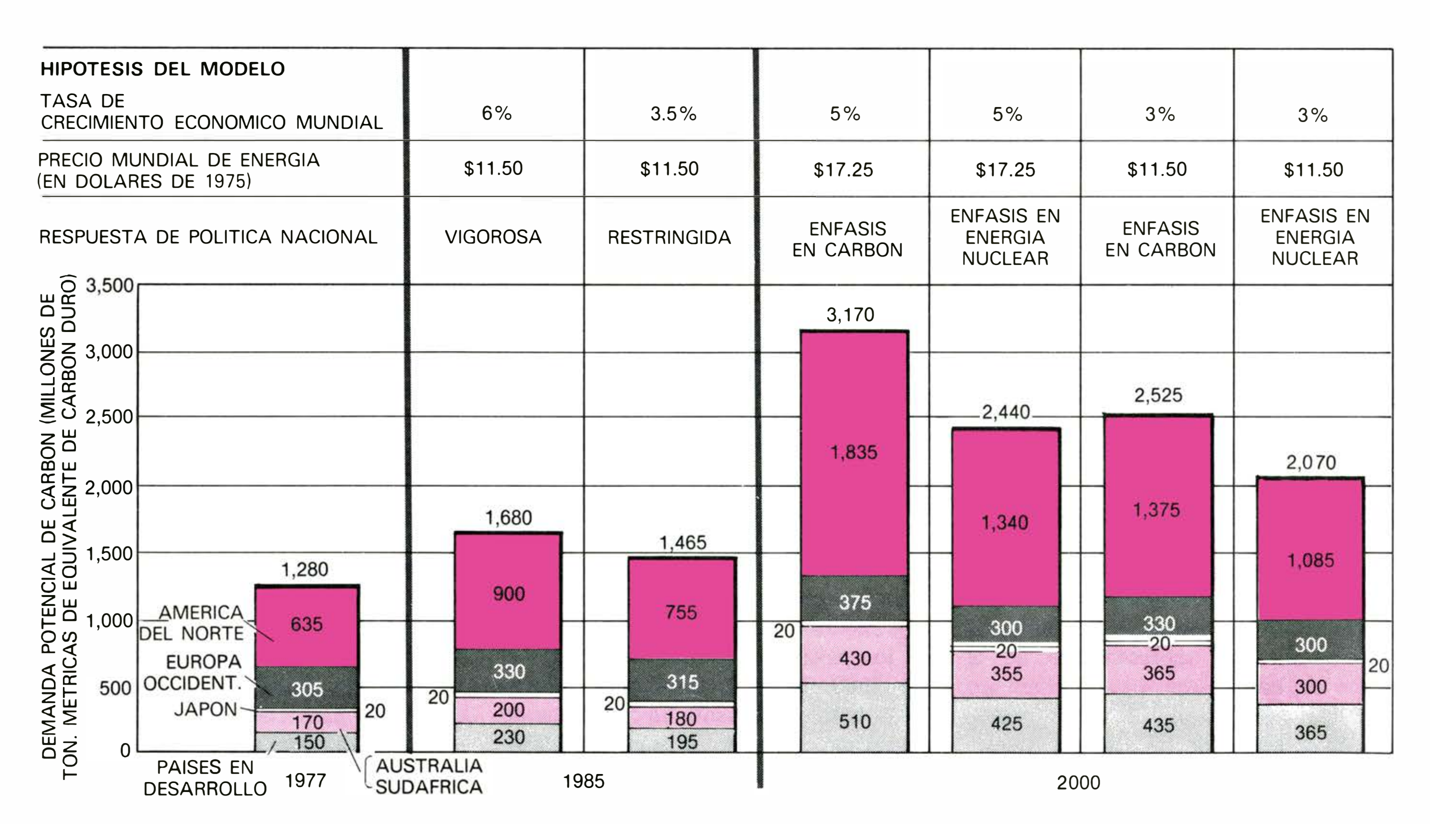
Al igual que en el caso de los cálculos previstos sobre la demanda, los valores más altos de la producción potencial de carbón se obtienen en el supuesto de un crecimiento económico mundial del 5 por ciento, de un precio de combustible igual a 17,25 dólares por barril de petróleo o su equivalente y de una política energética que ponga más énfasis en el carbón que en la energía nuclear. El volumen proyectado de carbón disponible, en estas condiciones, fue de 3170 millones de toneladas métricas, que representa un incremento de casi un 150 por ciento sobre la de 1977, que era de 1280 millones de toneladas. Se llega a dicho total proyectando máximos aumentos en América del Norte, en los países en vías de desarrollo, en Australia y en Sudáfrica. Se vio que América del Norte podía alcanzar una producción de 1835 millones de toneladas, es decir, casi tres veces el tonelaje de 1977. El proyecto más modesto de producción de carbón, en los países no comunistas y para el año 2000, fue de 2040 millones de toneladas, lo que representa un 59 por ciento de incremento sobre el total producido en 1977. Conforme ya hemos destacado, todos los proyectos de producciones potenciales de carbón, para el año 2000, exceden por amplios márgenes a las demandas proyectadas, bajo hipótesis comparables, de crecimiento económico, de precio de la energía y de énfasis político en combustibles.

Durante el período de 1985 a 2000, Europa Occidental y Japón serán netos importadores de carbón, potencialmente a gran escala, ya que su demanda excede a la producción local. Las regiones con notable potencial para exportar carbón serán América del Norte, regiones de América Latina, Sudáfrica, Asia meridional, Australia y algunos países comunistas. Entre los países en vías de desarrollo, que son pobres en petróleo, la India tendrá el máximo potencial de producción de carbón, posiblemente del orden de 200 millones a 250 millones de toneladas por año, para finales de siglo. Por lo que respecta a otros países en vías de desarrollo con reservas propias de carbón, ésta puede ser una alternativa doméstica para costear el petróleo importado y también una fuente posible de ingresos por exportaciones. Las importantes reservas de carbón que se encuentran en China y en la Unión Soviética pueden convertirlos en importantes exportadores de carbón.

Es dudosa la cantidad de carbón disponible para exportar a Europa y a Japón, porque depende de las estimaciones



DEMANDA POTENCIAL DE CARBON, para el año 2000. Se estimó en el Seminario sobre Estrategias Energéticas Alternativas, en el que participaron los autores del artículo. La demanda potencial de los países no comunistas, para 1985 y 2000, se compara con la de 1976, con diversas hipótesis de la tasa de crecimiento económico mundial, del precio de la energía en dólares constantes (1975) y de la eficacia y dirección de las políticas nacionales. En el año 1975, el precio mundial del petróleo era de 11,50 dólares por barril de crudo ligero de Arabia Saudita, embarcado en el Golfo Pérsico. La máxima demanda potencial para el año 2000 se calcula con una tasa de crecimiento económico del 5 por ciento anual, con un 50 por ciento de incremento en el precio real de la energía (a 17,25 dólares por barril de petróleo equivalente) y con unas medidas políticas que primen el carbón.



PRODUCCION POTENCIAL DE CARBON, para el año 2000. Se estimó también por los participantes en el Seminario, y se comparó con la producción de 1977. En todos los casos, excepto en uno (1985, con baja tasa de crecimiento económico, precios constantes de la energía y políticas restrictivas), la producción mundial de carbón excedía a la demanda potencial. El consumo de carbón en los países del bloque no comunista no debe restringirse, por falta de suministro, hasta el año 2000, si se exceptúan los cuellos de botella a corto plazo debidos a políticas energéticas limitativas. Para el año 2000, la producción potencial puede alcanzar 3170 millones de toneladas métricas, es decir, casi un 150 por ciento sobre la producción de 1977, si se supone un crecimiento económico elevado, un incremento del 50 por ciento en los precios de la energía y un énfasis nacional del carbón. La producción potencial de América del Norte representará un incremento de 190 por ciento con respecto a la de 1977. Para países en vías de desarrollo, los incrementos pueden ser del 200 por ciento. La productividad se define como el número de toneladas métricas de carbón por hombre y por día.

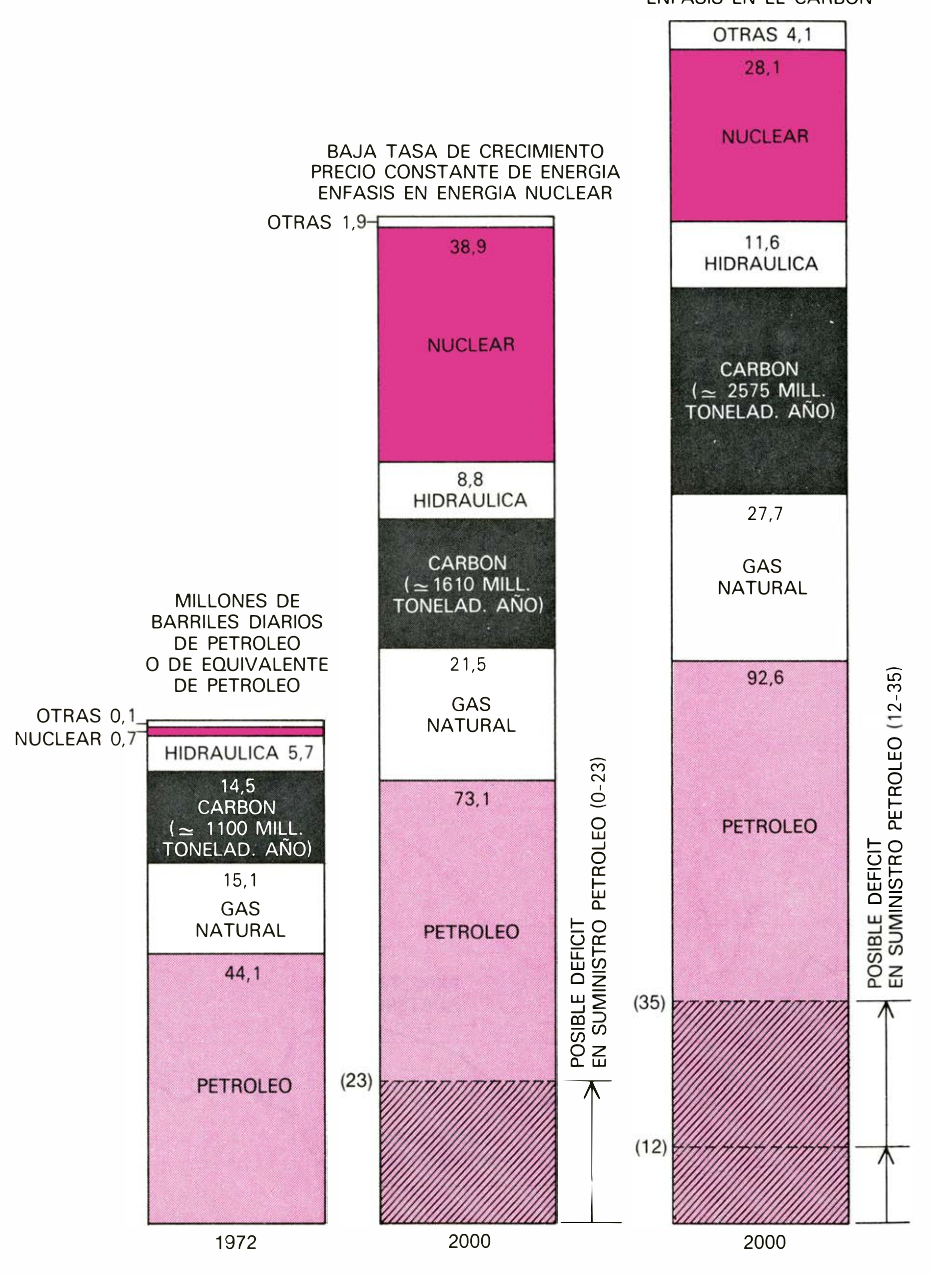
de los propios suministros interiores y de la demanda de cada una de las naciones exportadoras. No obstante, parece que la mayor fuente de energía, para los países industrializados energéticamente pobres, radicará en los embarques internacionales de carbón, que podrán llegar desde 400 millones hasta 900 millones de toneladas métricas, para el año 2000.

Los proyectos de suministro y de demanda, que se realizaron por el Seminario de Estrategias Energéticas Alternativas, incluyen tanto el carbón de vapor como el carbón siderúrgico. Los mayores incrementos disponibles para la exportación serán los de carbón de vapor. Sin embargo, a largo plazo, la distinción entre los diversos tipos de carbón se hará confusa, conforme la nueva tecnología haga que más tipos de carbón sirvan para fines siderúrgicos.

El carbón constituyó el accionamiento energético de la Revolución Industrial en Europa y en América del Norte, y también el principal combustible industrial y doméstico durante más de un siglo. Esta larga historia ha facilitado una compleja tecnología de producción de carbón. No obstante, la minería subterránea cuenta aún con excesiva mano de obra, sobre todo en Europa Occidental; por tanto, los costes de producción de carbón extraído en explotaciones subterráneas son muy sensibles a la inflación de salarios. La minería subterránea es además un trabajo arriesgado, hecho aprovechado por las presiones políticas y sociales para establecer y hacer cumplir unas severas reglamentaciones de seguridad e higiene, especiales para la minería del carbón. Como consecuencia, ha disminuido la productividad laboral en las minas, que se ha compensado, sólo en parte, con nuevo equipo y nuevas técnicas de minería.

Los aumentos significativos en la producción total de carbón deben obtenerse de expansiones adecuadas en minería de superficie, aunque en muchas regiones sigan siendo atractivas las explotaciones subterráneas de capas propicias de carbón de vapor y de capas relativamente delgadas de carbón de coque. La productividad, es decir, el número de toneladas de carbón por hombre y día, es típicamente dos o tres veces mayor en las explotaciones a cielo abierto que en las subterráneas. Sin embargo, las explotaciones a cielo abierto tienen sus propios problemas, especialmente las objeciones ambientales por las molestias causadas en extensas comarcas, durante las explotaciones mineras y antes de que se culmine el relleno de terrenos.

Los problemas ambientales, económicos y sociales relativos al carbón constituyen causas potenciales de conflicto en aquellas zonas del mundo que no cuenten con experiencia en minería de carbón a gran escala o en utilización del mismo. No está cerrada la cuestión de si la oposición ambiental y social constituirá un factor restrictivo en la expansión de la producción de carbón, fuera de los Estados Unidos. En cualquier caso, el éxito o el fracaso de los esfuerzos de Estados Unidos para negociar los problemas planteados y para alcanzar los objetivos del Presidente Carter de extender la utilización del carbón (1100 millones de toneladas métricas para 1985), pueden incidir en las actitudes de otros países con respecto a su uso.



DEMANDA TOTAL ENERGETICA proyectada para el año 2000, tal como se estimó en el Seminario; sugiere la alta cota que alcanzaría el consumo de petróleo, en los países no comunistas, si los suministros fuesen ilimitados. La reducida demanda energética proyectada con bajo crecimiento económico demuestra que las apetencias de petróleo sólo podrían satisfacerse, si se mantiene elevado el número de descubrimientos de campos petrolíferos, si la Organización de Países Exportadores de Petróleo (OPEP) no limita la producción, si el consumo de carbón se incrementa en un 50 por ciento y si se pone un fuerte énfasis en la energía nuclear. Se puede llegar a un déficit de 23 millones de barriles diarios, si el número de descubrimientos de nuevos campos petrolíferos disminuye y la OPEP restringe la producción. Con alto crecimiento económico, la alta demanda energética proyectada puede producir un déficit de petróleo entre 12 millones y 35 millones de barriles diarios. Si el carbón se lleva hasta su plena capacidad potencial de 3170 millones de toneladas métricas por año, con la demanda esperada de 2575 millones de toneladas, nos quedarían disponibles unos 600 millones de toneladas de carbón (equivalentes a unos 8 millones de barriles de petróleo por día) para compensar el déficit de petróleo.

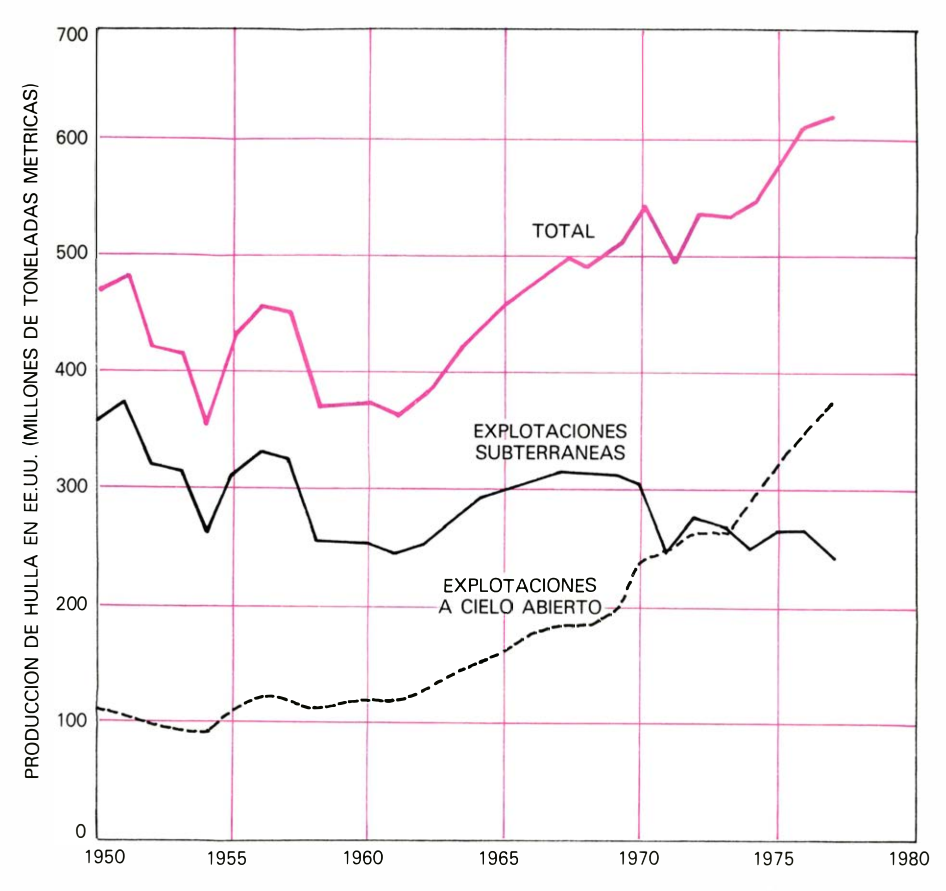
Como los Estados Unidos cuentan con la mitad de las reservas mundiales de carbón que están fuera del bloque comunista, y es el mayor productor y consumidor de carbón, su posibilidad de incrementar la producción importa, no sólo al mercado interior, sino también a futuros compradores de carbón en ultramar. Los proyectos elaborados por el Seminario indican que los Estados Unidos podrán alcanzar una producción anual, a finales del siglo, de 1100 a 2000 millones de toneladas métricas. La cifra menor corresponde a una estimación con precio real de energía constante (11,50 dólares por barril de petróleo o su equivalente) y una opción política con énfasis para la energía nuclear como sustituto principal del petróleo. La cifra mayor sale de un precio real de energía subido un 50 por ciento (a 17,25 dólares por barril de petróleo o equivalente) y de una política gubernamental de protección al carbón como principal combustible para sustituir al petróleo. En todos los casos, los proyectos reflejan incrementos máximos, sobre la producción de 1977, de 626 millones de toneladas métricas de carbón.

Aunque el 70 por ciento de las reservas de carbón de los Estados Unidos sólo pueden explotarse mediante minería subterránea, se espera que la mayoría de la nueva producción de carbón se obtenga por explotaciones a cielo abierto, durante los próximos 20 años, en los estados del oeste, especialmente al norte de Great Plains. Allí, las capas son potentes (de gran espesor), próximas a la superficie y bajo un terreno bastante horizontal. Los carbones del oeste poseen típicamente un menor poder calorífico que los carbones del este, pero en compensación muchos de aquéllos tienen un bajo contenido de azufre y, por tanto, emiten menos dióxido de azufre por tonelada de carbón quemado.

Una extensa minería a cielo abierto, en áreas vírgenes, contará con la oposición de los grupos ambientalistas y los intereses agrícolas locales. Hay dos objeciones principales para extender las explotaciones a cielo abierto en los estados del oeste: la potencial devastación de grandes eriales y los indeseables efectos económicos y sociales en una sociedad de pequeñas ciudades, fundamentalmente rural, de poca densidad de población, orientada agrícolamente, y de ranchos aislados. Los agricultores temen que algunos de los terrenos, que normalmente usa el ganado para pacer, deban arruinarse para futuros propósitos agrícolas, y que las necesidades de agua de la minería y los rellenos puedan pugnar con los intereses de los colonos. A los residentes locales les preocupa que el advenimiento masivo de trabajadores de la construcción y mineros rompa la estructura social de las pequeñas ciudades. Ponen reparos serios al ruido, al polvo y los accidentes que derivarían del gran número de trenes de carbón pasando por sus comunidades. Tales preocupaciones no pueden echarse en saco roto.

En el macizo del Río Powder, en los Estados de Wyoming y Montana, se están desarrollando explotaciones grandes a cielo abierto. Los dos estados han adoptado estrictas reglamentaciones para la minería a cielo abierto, estipulando que el desarrollo de la explotación debe ir pareja con el relleno, al objeto de evitar el tipo de devastación que se vio en las zonas montañosas de la minería del carbón en el este de los Estados Unidos. Adicionalmente, el Congreso ha elaborado una legislación para reglamentar las actividades de la minería a cielo abierto, a fin de proteger los recursos de agua y limitar la minería a ciertas áreas. En los próximos años, los mineros estarán en posición de demostrar que pueden llevar las operaciones mineras aceptablemente y que pueden rehabilitar de una manera eficaz los terrenos destrozados. Las compañías mineras son conscientes de que pueden hacerlo a un costo razonable con respecto al valor del carbón producido. Si la aceptabilidad de los programas de rehabilitación se demuestra de un modo concluyente, tal como ocurrió en Europa, las objeciones ambientalistas que ponen restricciones al futuro desarrollo de las minas de carbón deberán retirarse.

Para las áreas de la minería del carbón, se están elaborando planes también, a niveles federal y estatal norteamericano, al objeto de ayudar (y cubrir costos) a los gobiernos locales a través de servicios sociales y comunitarios debidamente ampliados. El temor a los ruidosos trenes de carbón deberá soslayarse con medidas que favorezcan el transporte del carbón por tubería, para llevarlo hasta el mercado, o bien mediante el desvío de algunas líneas de ferrocarril. No obstante, la continuada oposición a nuevas minas y los conflictos sobre las reglamentaciones detalladas de la minería, las jurisdicciones regula-

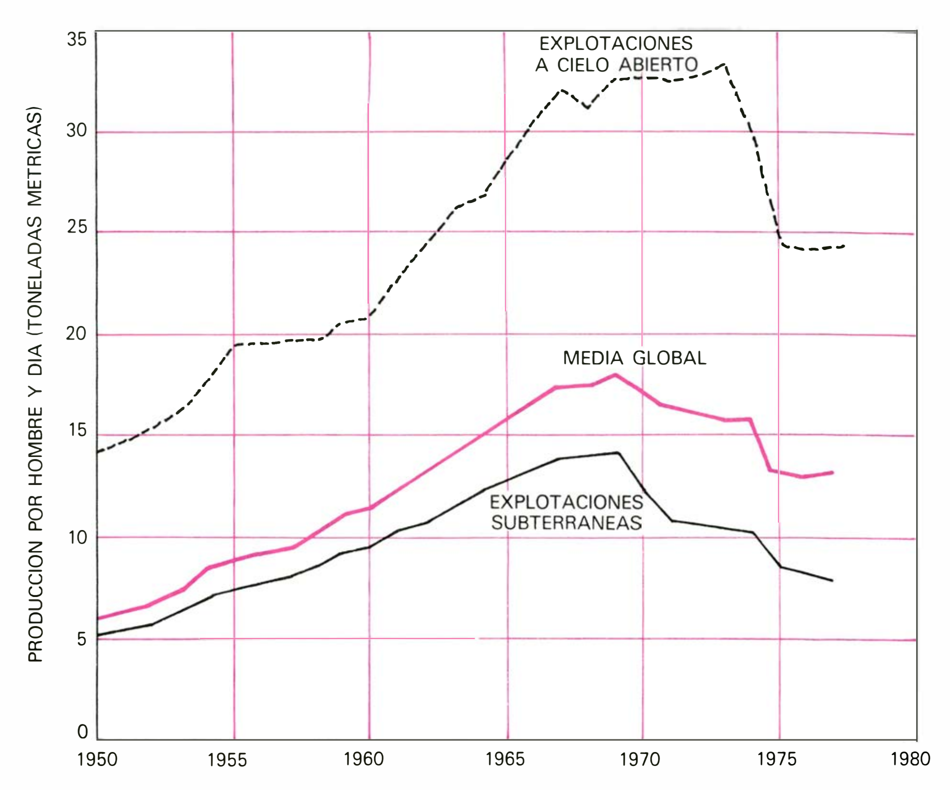


PRODUCCION TOTAL DE LAS MINAS DE CARBON, en los Estados Unidos. Desde 1954, fecha en que la producción descendió al mínimo tras la depresión de los años 1930, se incrementó menos del 3 por ciento anual. La producción de las explotaciones subterráneas está por debajo de los niveles del comienzo de la década de los años 1950. Pero, entre tanto, las explotaciones a cielo abierto (controvertidas por los ambientalistas) aumentaron en más del siete por ciento anual, durante los últimos diez años.

doras, las asignaciones de cánones, la planificación de la utilización de terrenos, los recursos de agua y la asistencia gubernamental local pueden frenar la expansión de actividades mineras a cielo abierto en los estados del oeste norteamericano. En conjunto, la incertidumbre es tan grande que lo proyectos de producción de carbón para el año 2000, elaborados por el Seminario, pueden diferir hasta en 500 millones de toneladas métricas anuales, según que las acciones gubernamentales primen o no, eficazmente, el desarrollo de los recursos de carbón de los Estados Unidos.

Los proyectos para el año 2000, que incorporan elevadas estimaciones de producción de carbón, suponen también incrementos sustanciales en las explotaciones subterráneas mineras de la parte este de los Estados Unidos, consecuentemente a los incrementos previstos en el precio de la energía. Desde hace algunos años, el costo del carbón extraído de explotaciones subterráneas han venido subiendo como consecuencia de las severas reglamentaciones de seguridad minera cada vez más numerosas, de una disminución en la productividad de las minas subterráneas y de un gradual agotamiento de las reservas subterráneas más accesibles. Es imposible predecir hasta cuándo continuará esta tendencia. El desarrollo del carbón subterráneo del este podría estimular, sin embargo, el empleo y la actividad económica en unas partes del país con antecedentes de desempleo y subdesarrollo. La continuada importancia de la minería del carbón en los estados del este se ha reflejado, recientemente, en algunas medidas políticas federales, que favorecen el carbón del este. En cualquier caso, la mayoría de los proyectos de producción de carbón de los Estados Unidos prevé notables incrementos no sólo en el oeste, sino también en el este y en el medio oeste.

La previsión de un elevado comercio internacional en carbón, hecha por el Seminario sobre Estrategias Energéticas Alternativas, implica un amplio incremento de las exportaciones de carbón por parte de los Estados Unidos. ¿Se dejará convencer Estados Unidos y aceptará las consecuencias ambientales y sociales y el agotamiento de un recurso nacional no renovable, que conlleva el suministro de energía a otros países? Se pueden adelantar buenos argumentos en favor de ello. Para los Estados Unidos, el carbón no constituye ningún recurso escaso; por tanto, al exportarlo para ayudar a otros países a cubrir sus necesidades energéticas se compensa simultáneamente el elevado coste de importar petróleo, lo cual podría parecer de un claro interés nacional. La exportación de carbón tendría un efecto positivo sobre el desarrollo económico, el empleo y la balanza de pagos del comercio exterior. Sin embargo, no se darán estos beneficios a no ser que se cumplan tres condiciones. Primera: la demanda de exportaciones de carbón debe asegurarse en firme, por contratos a largo plazo, basados en la esperanza del comprador de que el carbón estará disponible para la exportación. Segunda: las consecuencias ambientales y sociales de unas exportaciones ampliadas de carbón deben considerarse aceptables para el pueblo norteamericano. Tercera: tiene que prepararse una planificación con suficiente antelación para suministrar todos los medios necesarios que requerirá un aumento de las exportaciones de carbón, en particular la ampliación de puertos de gran calado. Estos puntos no son todavía materias centrales en el debate energético de los Estados Unidos. Convendría que lo fueran en los próximos años.



PRODUCTIVIDAD EN LAS MINAS DE CARBON de los Estados Unidos, medida en toneladas por hombre y día; aumentó rápidamente entre 1950 y 1969. El nivel alcanzado disminuyó desde entonces, por una serie de causas, entre las que se contaban las nuevas regulaciones de seguridad e higiene mineras y la legislación, en caso de explotaciones a cielo abierto, que obligaba a la restauración de todo paraje afectado por las labores mineras. La productividad en las minas de carbones duros (antracitas y hullas) de Europa, casi todas ellas con labores subterráneas, es mucho menor que en las de los Estados Unidos; unas 2,5 toneladas por hombre y día, en Gran Bretaña; aproximadamente de 1,9 toneladas en Francia y alrededor de 3,8 toneladas en Alemania Occidental. (Los lignitos son carbones blandos.)

Grandes incrementos en la producción de carbón exigen la construcción y la explotación de extensas minas y redes de transporte para obtener el carbón y llevarlo hasta sus mercados. Por el Seminario se prepararon análisis especiales de equipos mineros, de medios de transporte, de mano de obra y de otros recursos, para una rápida expansión de la producción de carbón en los Estados Unidos. Un estudio financiado por la Bechtel Power Corporation demostró que, para alcanzar una producción anual de 2000 millones de toneladas métricas de carbón, a finales del siglo, Estados Unidos necesitaría abrir: 377 nuevas minas de explotación subterránea en el este, cada una de las cuales debía dar 2 millones de toneladas al año; 75 minas de explotación a cielo abierto, en el este, cada una para 4 millones de toneladas al año; 232 minas de explotación a cielo abierto, en el oeste, dando 6 millones de toneladas al año; y 8 minas de explotación subterránea en el oeste, dando dos millones de toneladas al año. El coste total en dólares de 1975 sería de 32.000 millones. La inversión en medios de transporte para llevar el carbón a los consumidores sería mucho más elevada: 86.000 millones de dólares. Esta inversión facilitaría 1400 trenes "unificados" (composiciones puras de unos 100 vagones-tolva, cada tren, exclusivamente de-

dicados al transporte de carbón entre bocamina y el consumidor), 3200 trenes convencionales, 500 gabarras fluviales de gran capacidad, 9400 camiones y nueve líneas de tubería de transporte de carbón, con una capacidad de 25 millones de toneladas al año cada una.

La necesaria capacidad de transporte para el carbón a granel puede facilitarse por ferrocarril, por tubería o por una combinación de ambos medios. El ferrocarril es más conveniente cuando se requiere máxima flexibilidad para transportes locales y para volúmenes de pequeños a moderados. El transporte por tubería será más económico para volúmenes grandes y constantes a distancias largas y recorridos fijos. Actualmente ya existe una amplia red ferroviaria. El desarrollo de las tuberías de transporte de carbón en los Estados Unidos, por el contrario, tropiezan con enormes obstáculos políticos, que tendrán que resolverse. Se ha incoado una legislación federal, pero su tramitación y aprobación no está asegurada.

El cuello de botella más importante, para desarrollar una creciente exportación de carbón, está en las instalaciones portuarias marinas. Aunque los Estados Unidos es ya el primer exportador del mundo en carbón (50 millones de toneladas métricas en 1977), sus instalaciones portuarias resultan, en general, inadecuadas para cualquier aumento notable de las exportaciones de carbón de vapor y tienen una capacidad limitada para trabajar con grandes barcos de carga a granel.

El Seminario concluyó que las necesidades de infraestructura y transporte, para una gran expansión del carbón, eran importantes pero no inalcanzables, en vista de que la tasa proyectada de crecimiento máximo anual, para la producción de carbón en los Estados Unidos, es de un 6 por ciento. No obstante, incluso objetivos más modestos llegarán a ser difíciles de conseguir, si el grueso de las inversiones se pospone hasta mediada la próxima década. Las industrias que facilitan los recursos y los equipos mineros y de transporte son capaces de cumplimentar su parte. Pero se precisan señales claras de una mayor demanda, para poner en marcha las ampliaciones necesarias.

Incluso con un supuesto incremento del 50 por ciento en el precio real de la energía (17,25 dólares por barril de petróleo o su equivalente en 1975) y con políticas protectoras del desarrollo de combustibles sintéticos, los estudios del Seminario indican sólo unos modestos potenciales, para el año 2000, de la conversión del carbón en gas y petróleo sintéticos. El máximo volumen de combustibles sintéticos basados en el carbón, según nuestros estudios, era ligeramente superior a los 4 millones de barriles diarios de petróleo o su equivalente, produciéndose en los Estados Unidos las dos terceras partes de los gases sintéticos y la mayor porción de los petróleos sintéticos. Con un típico rendimiento de conversión del 60 a 70 por ciento, este volumen de combustible sintético requeriría unos 500 millones de toneladas métricas de carbón.

Para los Estados Unidos, con sus grandes reservas de carbón, la producción de combustibles sintéticos a partir del carbón puede constituir una alternativa atractiva ante las crecientes importaciones de petróleo y de gas natural licuado. Esto permitiría a los usuarios mantener los equipos existentes de combustión de petróleo y de gas, colocando así la carga de la nueva inversión sobre los productores y transformadores de la energía, más que sobre los distribuidores y consumidores. Esto simplificaría también el control de la polución del aire, al concentrar la utilización del carbón en pocas ubicaciones. Por otro lado, la conversión del carbón en otras formas de combustible se añade al coste energético del usuario, habida cuenta del relativamente bajo rendimiento de la conversión y de los elevados costes de los equipos de conversión. La producción de combustible sintético requiere la explotación de volúmenes de carbón mayores que cuando se utiliza directamente el carbón, y si tiene lugar en el oeste de los Estados Unidos, se precisaría una carga adicional en los recursos de agua para refrigeración y para relleno tras la explotación a cielo abierto.

A no ser que haya fuertes aumentos en el precio del petróleo, los combustibles sintéticos continuarán costando más que el petróleo o el gas naturales. Debido al elevado coste de los combustibles sintéticos, éstos no podrán desarrollarse sin una política vigorosa de soporte por parte de los gobiernos. Cabe pensar en que la disminución de la producción mundial de petróleo y la elevación de sus precios determine que algunos países que son grandes importadores de petróleo (en particular Estados Unidos y posiblemente Europa Occidental) comiencen a explotar sus recursos de carbón para producir petróleo y gas sintéticos. Cuando se dispone de una capacidad de combustible sintético, se exige una acción política anticipada que financie la puesta a punto de la tecnolo-

PRODUCCION DE CARBON en los Estados Unidos, para el año 2000. Puede elevarse hasta más de 2 millones de toneladas métricas anuales, si las tasas de crecimiento económico son altas, si los precios mundiales de la energía suben el 50 por ciento sobre los actuales y si las políticas gubernamentales protegen la producción y utilización del carbón. La curva inferior muestra que la producción potencial del carbón puede ser notablemente menor, si las tasas de crecimiento económico son bajas y si las políticas favorecen más a la energía nuclear que al carbón. Los porcentajes adyacentes a los segmentos curvos indican las tasas medias de crecimiento anual en la producción de carbón, en los dos modelos.

gía del combustible sintético comercial. Los tiempos aludidos son largos, y los gobiernos se mostrarían prudentes si no esperasen a la subida de los precios del petróleo para conceder incentivos a las inversiones en combustible sintético.

Se han propuesto ya varios grandes proyectos de combustible sintético en los Estados Unidos, pero no serán viables sin el apoyo del Gobierno, en forma de garantías de empréstitos, disposiciones de precios especiales o un generoso tratamiento regulador. Aunque algunos proyectos obtuvieron la aprobación provisional de los organismos competentes, el nivel adecuado de soporte federal, para los combustibles sintéticos, todavía está debatiéndose en el Departamento de Energía y en el Congreso. Si la decisión del Gobierno fuese favorable en corto plazo, los combustibles sintéticos podrían comenzar a comercializarse antes de 1985 y alcanzarían niveles significativos en el año 2000.

El carbón es el combustible fósil que probablemente se mantendrá abundante y de bajo costo durante el resto de este siglo y en buena parte del próximo. Por eso constituye el principal de los combustibles de sustitución para atender las necesidades energéticas del período transitorio entre la era de la abundancia de petróleo y gas, en decadencia, y la futura era para la que el mundo deberá desarrollar nuevas fuentes alternativas de energía. El carbón será incluso más importante si las naciones se muestran reacias, como hay algunas hoy día, a volcarse vigorosamente en la energía nuclear. Inversamente, las decisiones que impidan o limiten el consumo de carbón presionarán en favor de un mayor desarrollo de la energía nuclear y en favor de unas medidas restrictivas de conservación energética, como alternativa a una demanda de petróleo insatisfecha.

Si el carbón tiene que servir de puente hacia una nueva era energética, los gobiernos deberán adoptar rápidamente las medidas políticas precisas para permitir y promover una utilización intensificada. Habrá que recurrir a fuertes inversiones, públicas y privadas, para la apertura de nuevas minas, la construcción de los sistemas de transporte, incluyendo instalaciones de puertos de gran calado, y el desarrollo de la infraestructura para manipular y quemar el carbón por los usuarios. Finalmente, la explotación intensificada del carbón exige soluciones satisfactorias para los problemas ambientales asociados con la minería y la combustión del carbón, dondequiera que sea.

La formación de membranas celulares

Las dos caras de una membrana biológica difieren en su estructura y función. Estudios sobre virus animales y bacterias han contribuido a revelar la forma como se mantiene esta asimetría durante el crecimiento de la membrana

Harvey F. Lodish y James F. Rothman

La membrana que rodea una célula viva es algo más que una simple envuelta o receptáculo; no sólo define la extensión de la célula sino que actúa manteniendo una diferencia entre el interior y el exterior. Por ejemplo, algunos iones son bombeados hacia el interior de la célula mediante grandes moléculas incrustadas en la membrana, mientras que otros sufren el efecto contrario: son bombeados hacia fuera. La célula toma los nutrientes y los concentra en el interior gracias a la acción de otros componentes de la membrana.

Para que una membrana pueda mantener tales gradientes de concentración se precisa, como exigencia absoluta, que forme una vesícula cerrada; en caso contrario, una sustancia bombeada a través de la membrana saldría inmediatamente fuera. Todas las membranas biológicas conocidas forman compartimentos cerrados. Otra propiedad esencial intrínseca a la membrana estriba en su lateralidad: la superficie interior y la superficie exterior deben cumplir funciones distintas.

Si no fuera así, un ion o molécula bombeada hacia el interior en un punto podría ser impulsado hacia fuera en otro, con el derroche energético correspondiente. Este carácter de lateralidad se manifiesta asimismo en otras propiedades de las membranas. Por ejemplo, en las células animales los receptores de hormonas y de otros agentes químicos están asociados a la membrana, y en este sentido son marcadores que identifican una célula ante sus vecinas. Estos elementos de comunicación intercelular deben ser accesibles en la superficie exterior; en la interior serían inútiles.

Hace ya algunos años, quedó establecido que la asimetría funcional de las membranas refleja una asimetría estructural subyacente. Las moléculas proteicas incrustadas en la membrana o ligadas a ella ofrecen una orientación fija. Unas se presentan sólo en la superficie interior, otras sólo en la exterior. Otras, por fin, se extienden por todo el espesor de la membrana, pero incluso éstas pueden observarse invariablemente en orientación fija y asimétrica. Todas las proteínas de un mismo tipo se orientan en la misma dirección. Se ha encontrado que las moléculas de fosfolípidos, más pequeñas, que constituyen la matriz estructural de las membranas, presentan, también, una distribución asimétrica; si bien en éstas la asimetría es parcial, no absoluta.

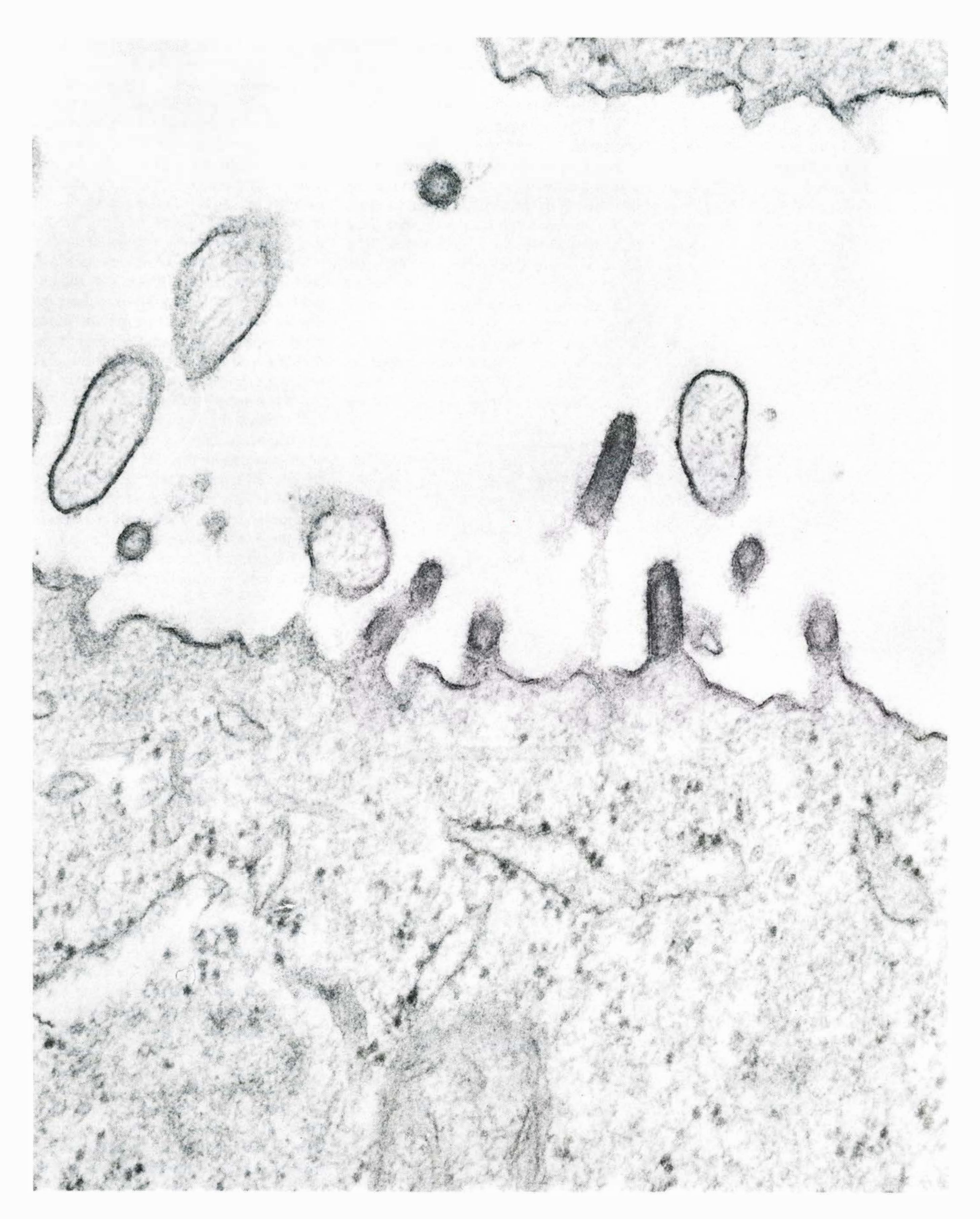
Los biólogos tienen un reto en la comprensión de cómo las células vivas se replican y dividen; y una parte importante de este problema es cómo forman membranas las células. Cuando una célula se divide, los componentes de la nueva membrana deben sintetizarse y ensamblarse de un modo conveniente. La nueva membrana debe situarse con gran precisión; en particular, los nuevos componentes han de presentar una distribución topográfica idéntica a la de los antiguos componentes establecidos, para asegurar de este modo el correcto funcionamiento de la membrana recién formada.

Las grandes líneas de los modelos celulares empleados para resolver estos problemas se han clarificado mucho recientemente, aunque muchos detalles bioquímicos permanecen aún confusos. Este progreso ha sido posible gracias al estudio de unos pocos tipos de membranas altamente especializadas y que incorporan un pequeño número de componentes moleculares que resultan idóneos para la investigación. Sobre esta base experimental hemos planteado la presente discusión.

Las células más sencillas son las bacterias. Tienen sólo una membrana, la membrana plasmástica que envuelve la célula. Estos organismos son procariotas, es decir, células sin un núcleo bien definido. Las plantas superiores y los animales están formados por células eucariotas, dotadas de verdadero núcleo. Además de la membrana citoplasmática, los eucariotas presentan una gran diversidad de membranas internas especializadas que rodean y definen las estructuras subcelulares llamadas orgánulos. El propio núcleo se encuentra rodeado por un sistema de membranas. En las mitocondrias hay dos membranas; la interior tiene un papel esencial en la fabricación de adenosín trifosfato (ATP), la "moneda energética", que se gasta en el interior de la célula. Los cloroplastos de las células de las plantas verdes presentan también dos membranas, que pueden operar de manera semejante. Otro orgánulo de particular importancia en el estudio de las membranas es el retículo endoplasmático, especializado en la fabricación de proteínas de secreción y materiales de la nueva membrana.

En nuestro trabajo nos hemos interesado principalmente en la membrana citoplasmática. (No debe confundirse con la pared celular que envuelve, por fuera de la membrana, a las células bacterianas y vegetales. No se trata de ninguna membrana, sino de un órgano esquelético poroso.) Todas las células tienen una membrana citoplasmática, y en los animales es la única barrera que separa la célula de su ambiente. Existen interesantes diferencias entre la membrana citoplasmática y algunas membranas celulares internas de las células eucariotas, pero todas las membranas son considerablemente similares en sus estructuras básicas principales.

Si se separa una membrana en sus



REGION DE LA MEMBRANA CELULAR, que envuelve la partícula vírica cuando el virus abandona la célula infectada. La membrana que se apropia el virus incorpora un solo tipo de proteína superficial, lo que hace de la cubierta del virus un sistema experimental muy apropiado para estudios sobre la formación de membranas. Cada molécula de proteína superficial se orienta en la membrana formando una espícula sobre la superficie del virus. La célula está infectada por el virus de la estomatitis vesicular (VSV), que provoca alteraciones similares a la gripe en animales de granja. Las partículas víricas son los pequeños objetos obscuros, en forma de cigarro puro, que emergen en el espacio intercelular de la parte superior; algunos se observan lateralmente, por lo que parecen esféricos. Las vesículas amplias e irregulares son material celular aislado al seccionar las células. Esta micrografía electrónica, que aumenta la imagen unas 60.000 veces, fue realizada por David Knipe, del Instituto de Tecnología de Massachusetts.

componentes moleculares, se encuentra que está formada por tres tipos de sustancias: lípidos, proteínas y glúcidos. Las moléculas lipídicas son, con diferencia, las más abundantes; las proteínas, sin embargo, son moléculas mucho mayores. En definitiva, estos dos componentes moleculares se hallan equilibrados respecto a la masa total. Los componentes lipídicos de la membrana están representados por escasos modelos moleculares repetitivos; en cambio, hay un rico surtido de tipos proteicos, por más que algunos sólo estén representados por unas cuantas moléculas por membrana. Los glúcidos constituyen el componente minoritario (en masa) y están asociados principalmente a la membrana plasmática, más que a membranas internas. Los carbohidratos están en la membrana siempre combinados químicamente con otros componentes; se unen a los lípidos para formar glucolípidos y a las proteínas para formar glucoproteínas.

Los lípidos son responsables de la integridad estructural de la membrana. Se sabe que cada molécula lipídica es anfipática, lo cual significa que tiene una parte hidrofóbica y otra hidrofílica. Si la parte hidrofóbica se separa del resto de la molécula, será insoluble en agua, igual que lo es el aceite. Si se aislara la parte hidrofílica, sería insoluble en aceite pero soluble en agua.

Las membranas incluyen dos tipos de lípidos: colesterol y fosfolípidos. El colesterol aparece casi exclusivamente en la membrana plasmática de las células de los mamíferos y está ausente por completo en las bacterias. La cabeza hidrofílica del colesterol es el grupo hidroxilo (OH). Los fosfolípidos se encuentran en todas las membranas. La región hidrofílica incluye un grupo fosfato cargado negativamente (PO_4^-); en muchos fosfolípidos, otra entidad química, tal como un grupo amino (NH_3^+), aporta una carga positiva de compensación.

Dado que las dos partes de una molécula lipídica tienen solubilidades incompatibles, las moléculas se organizan espontáneamente en forma de un doble nivel o doble capa. Así, la porción hidrofóbica de la molécula está protegida del agua, mientras que las cabezas hidrofílicas están inmersas en ella. Por tanto, una doble capa laminar es la configuración de mínima energía para una suspensión de lípidos. En el único punto en que las colas no polares deben interaccionar con el agua es en el borde de la lámina, e incluso este contacto desfavorable puede evitarse mediante un simple plegamiento de la doble capa para formar una vesícula cerrada que no presenta bordes.

Los lípidos de todas las membranas presentan una organización similar, con dos superficies hidrófilas separadas por una lámina interna hidrófoba. Esta doble capa lipídica determina la morfología de las membranas y la formación de láminas o vesículas. La estructura de la doble capa determina, asimismo, una de las funciones esenciales de la membrana: su impermeabilidad a las moléculas solubles en agua, ya que son insolubles en la región central grasa.

FOSFATIDIL ETANOLAMINA

AMINO

FOSFATO

GLICEROL

PARTE HIDROFILICA (CABEZA)

PARTE HIDROFOBICA (COLA)

COLESTEROL

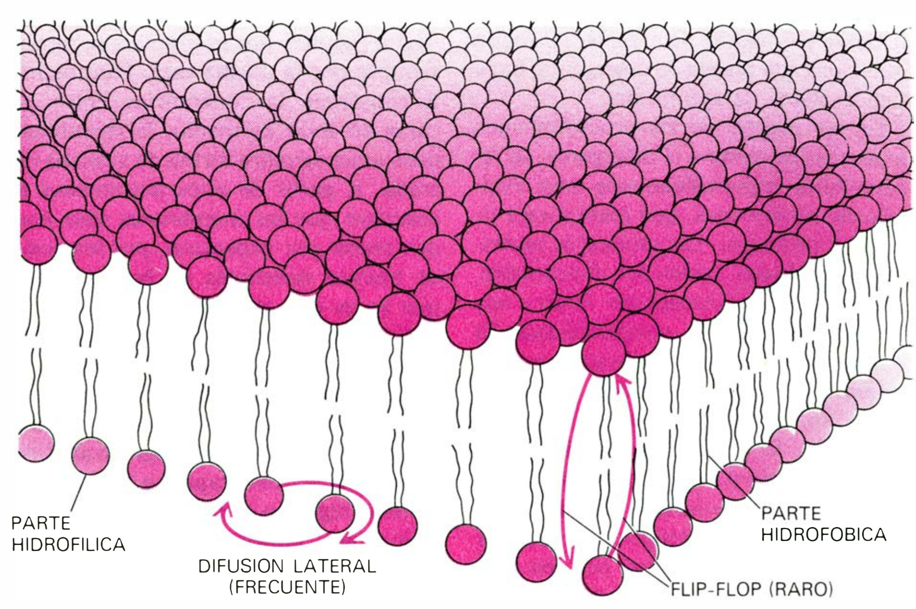
MOLECULAS LIPIDICAS de la membrana celular: comprenden fosfolípidos y colesterol. Ambos tipos de lípidos son anfipáticos, es decir, un extremo de cada molécula, llamado grupo de cabeza, es polar e hidrofílico; si se aislara, este grupo sería soluble en agua. El otro extremo, la cola, es no polar e hidrofóbico, o insoluble en agua. El colesterol se localiza en el exterior de la membrana plasmática de las células de mamíferos; fosfolípidos de varios tipos son componentes de todas las membranas.

La doble capa no permanece estática; al contrario, cada nivel por separado es un fluido bidimensional. Las moléculas lipídicas se encuentran libres para difundir lateralmente de un modo similar a las moléculas de una fina película de líquido. Cambian de posición con una frecuencia del orden de un millón de veces por segundo. Sin embargo, en la tercera dimensión, la movilidad de los lípidos está severamente restringida. Para que una molécula de lípido salte de un nivel a otro, fenómeno llamado "flip-flop" (sistema biestable), la cabeza polar debe pasar a través de la parte central hidrofóbica de la membrana, en la cual es insoluble. Recientes medidas indican que la tasa de "flip-flop" es tan baja que una molécula lipídica determinada hace un paso no más de una vez al mes.

Dado que la doble capa lipídica es un fluido bidimensional, cualquier proteína incluida en él puede difundir lateral-

mente. El carácter fluido simplifica la tarea del ensamblaje de la membrana, ya que ambas, las moléculas lipídicas y las proteínas, pueden insertarse en cualquier parte de un nivel con la seguridad de que eventualmente permanecerán en otro punto. Al mismo tiempo, la baja frecuencia de "flip-flop" permitiría a los dos niveles opuestos conservar sus distintas composiciones de lípidos y proteínas. La doble capa lipídica ha resultado ser de composición asimétrica en todas las membranas biológicas examinadas a este respecto. Sin embargo, la función de la asimetría está mal establecida.

En las proteínas, las interacciones de los grupos químicos hidrofílicos e hidrofóbicos no son menos importantes que en los lípidos. Las proteínas son polímeros (polipéptidos) formados por aminoácidos unidos en una secuencia lineal. De los veinte aminoácidos especificados por el código genético, seis son fuertemente hidrofóbicos y unos pocos débilmente hidrofóbicos; el resto son hidrofílicos. Si consideramos la proteína como mera cadena lineal de aminoácidos, costará distinguir entre las unidades hidrofóbicas e hidrofílicas de la secuencia. Sin embargo, la conformación inicial de las moléculas proteicas no es la de una simple cadena sino la de una estructura densamente plegada. De este modo, las proteínas solubles en el citoplasma celular poseen abundantes unidades hidrofílicas en su superficie molecular y en cambio su zona interna excede en unidades hidrofóbicas.



DOBLE CAPA DE FOSFOLIPIDOS, que forma la matriz estructural de una membrana. Los lípidos están dispuestos cola frente a cola, de modo que sólo las cabezas hidrofílicas queden expuestas al medio acuoso en ambas caras de la membrana. Esta es la configuración de mínima energía para una suspensión de lípidos en agua. Cada capa sencilla constituye un fluido bidimensional: las moléculas lipídicas pueden difundir lateralmente como una película líquida, pero raramente pueden ejecutar una transición "flip-flop" de una capa a otra. Los lípidos son responsables de la integridad estructural.

Las proteínas asociadas a las membranas se dividen en dos clases: constitutivas y periféricas. Las proteínas constitutivas tienen una porción molecular incluida en la doble capa lipídica. Se ha encontrado que todas las proteínas constitutivas que se han estudiado detalladamente abarcan toda la anchura de la doble capa, y poseen asimismo regiones que sobresalen a ambos lados de la membrana. La otra clase, formada por proteínas periféricas de la membrana, no están insertadas en la doble capa, y todas ellas se sitúan en una u otra superficie. Cada molécula de proteína periférica está ligada a una proteína constitutiva.

En las proteínas periféricas, el balance de aminoácidos hidrofílicos e hidrofóbicos es muy similar al de la proteínas citoplasmáticas. Las proteínas periféricas pueden separarse de la membrana por procedimientos que no destruyan la integridad de la doble capa lipídica; una vez separadas de la misma, son solubles en agua. Por el contrario, las proteínas constitutivas son generalmente insolubles en agua porque comprenden regiones importantes en las que la mayoría de los aminoácidos expuestos son hidrofóbicos. Estas regiones son las que están incluidas en el núcleo hidrofóbico de la membrana y, por tanto, no pueden separarse sin la rotura de la doble capa de lípidos.

Las regiones de una proteína constitutiva que se extienden en el fluido externo o en el citoplasma tienen el carácter hidrofílico común a las proteínas solubles. Estas regiones hidrofílicas evitan esencialmente la posibilidad de una transición "flip-flop" de las proteínas de la membrana. Si la pequeña cabeza polar de una molécula lipídica raramente puede pasar a través de la parte central de la doble capa, con mayor razón la posibilidad de que lo haga una molécula proteica, cuyo volumen es muy superior, queda permanentemente excluida.

El glóbulo rojo es una célula adecuada para investigar la topografía de la membrana, aunque no para estudios de ensamblaje. La membrana plasmática del glóbulo rojo posee dos especies de proteínas constitutivas y varias de tipo periférico; su orientación puede determinarse fácilmente. A partir de estos estudios de los glóbulos rojos, se han podido enunciar algunos principios sobre la construcción de la membrana. Todas las proteínas constitutivas se hallan insertadas asimétricamente, de modo que cada molécula de las de un determinado tipo de proteína tienen una orientación idéntica, en la doble capa lipídica. La doble capa lipídica es, a su vez, asimétrica. Los carbohidratos pueden ligarse a las proteínas o a los lípidos, pero en todos los casos se encuentran invariablemente en la cara exterior de la membrana, faltando de la superficie que contacta con el citoplasma. Las proteínas periféricas suelen ubicarse en la cara citoplasmática.

En una suspensión de lípidos pueden obtenerse de un modo espontáneo vesículas cerradas. Y en principio no es descabellado suponer que una membrana biológica completa y funcional pudiera formarse de manera similar. Ciertos virus, fibras musculares y ribosomas se autoensamblan de una forma espontánea, pero estas estructuras poseen una elevada organización incluso de tipo cristalino; en cambio no ocurre así a nivel de las membranas que estamos considerando.

La hipótesis del autoensamblaje puede asimismo comprobarse. Una membrana puede dispersarse en moléculas proteicas y moléculas lipídicas individuales rompiendo la doble capa con una elevada concentración de detergente. El detergente, lípido sintético que no forma una doble capa, origina unas gotículas llamadas micelas. Los lípidos y las proteínas de la membrana no necesitan permanecer juntos para eludir el agua, ya que se encuentran protegidos por las micelas del detergente. Los componentes de una membrana pueden se-

pararse con una concentración suficientemente elevada de detergente, de suerte que, por cada micela de detergente, haya una proteína de membrana.

Si por alguno de los procedimientos existentes se elimina el detergente, se reconstituirá espontáneamente una vesícula de membranas, que incorporará lípidos y proteínas. Las moléculas proteicas individuales suelen ser funcionales, capacitadas para transportar iones a través de barreras de permeabilidad. El único error en estas membranas reconstruidas suele presentarse con la pérdida de la asimetría. Las proteínas están incluidas en la doble capa lipídica, pero ya no se presentan en la misma dirección todas las moléculas de un determinado tipo. Hay excepciones que son interesantes por sí mismas, pero en muchos casos desaparece la asimetría absoluta de la membrana original.

¿Por qué en estas experiencias no se reconstituye la asimetría junto con el resto de la estructura de la membrana? En el autoensamblaje de una membrana, tiene lugar simultáneamente la formación de la doble capa de lípidos y la inserción de las proteínas de la membrana. Al no haber, durante el ensamblaje, barrera de permeabilidad, las proteínas pueden insertarse en ambas caras; así resulta una estructura aproximadamente simétrica.

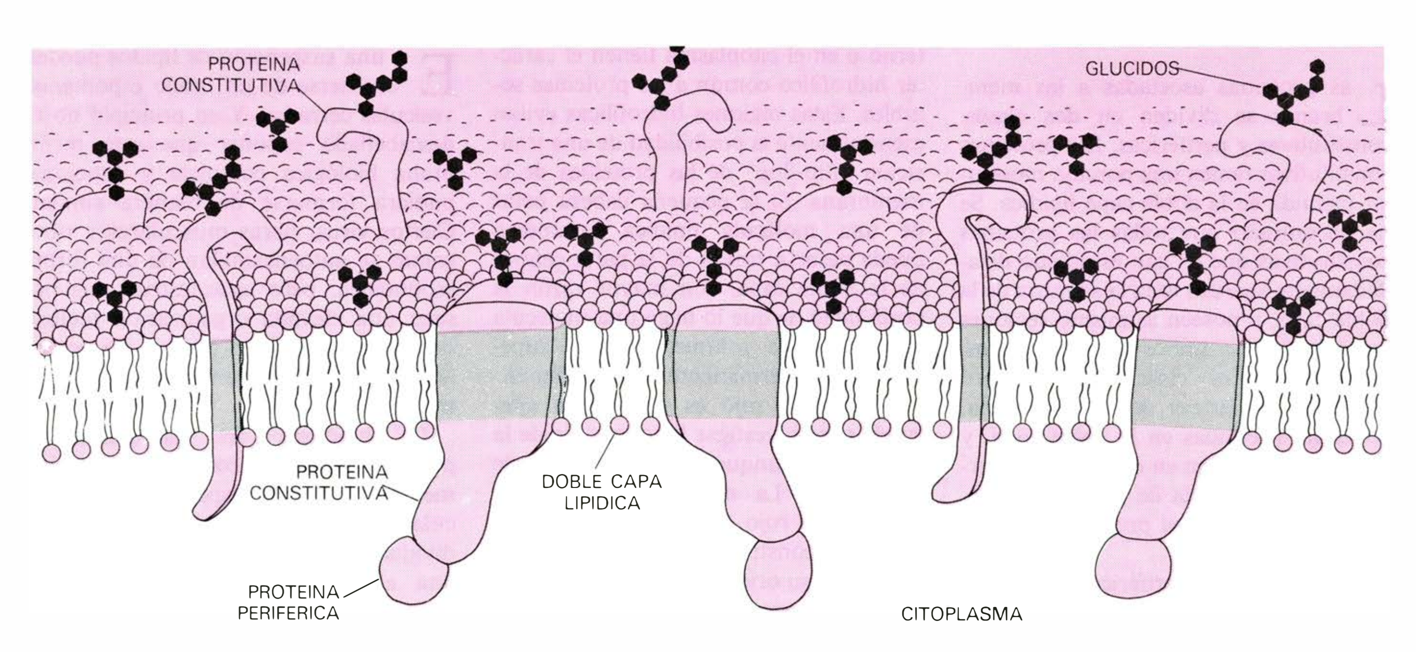
La hipótesis alternativa a la del autoensamblaje es que las proteínas constitutivas de la membrana deban su orientación al tiempo de inserción. Este mecanismo requiere que la proteína se inserte siempre por la misma cara. Para definir y conservar las diferencias entre las caras en todo momento, la membrana debe tener siempre forma de vesícula cerrada. A partir de estos argumentos se concluye que las membranas sólo pueden crecer por expansión, por inserción de nuevos materiales en la membrana que ya es una vesícula cerrada.

Este mecanismo presenta también una dificultad: si una proteína (o un lípido) se inserta en una cara de la membrana, en muchos casos deberá ser impulsada una región hidrófila a través del interior de la doble capa. La incapacidad de estas regiones para atravesar la membrana hace posible la asimetría de las proteínas que se han insertado. Nuestros experimentos han ayudado a descubrir los métodos por los cuales esta estrategia es viable.

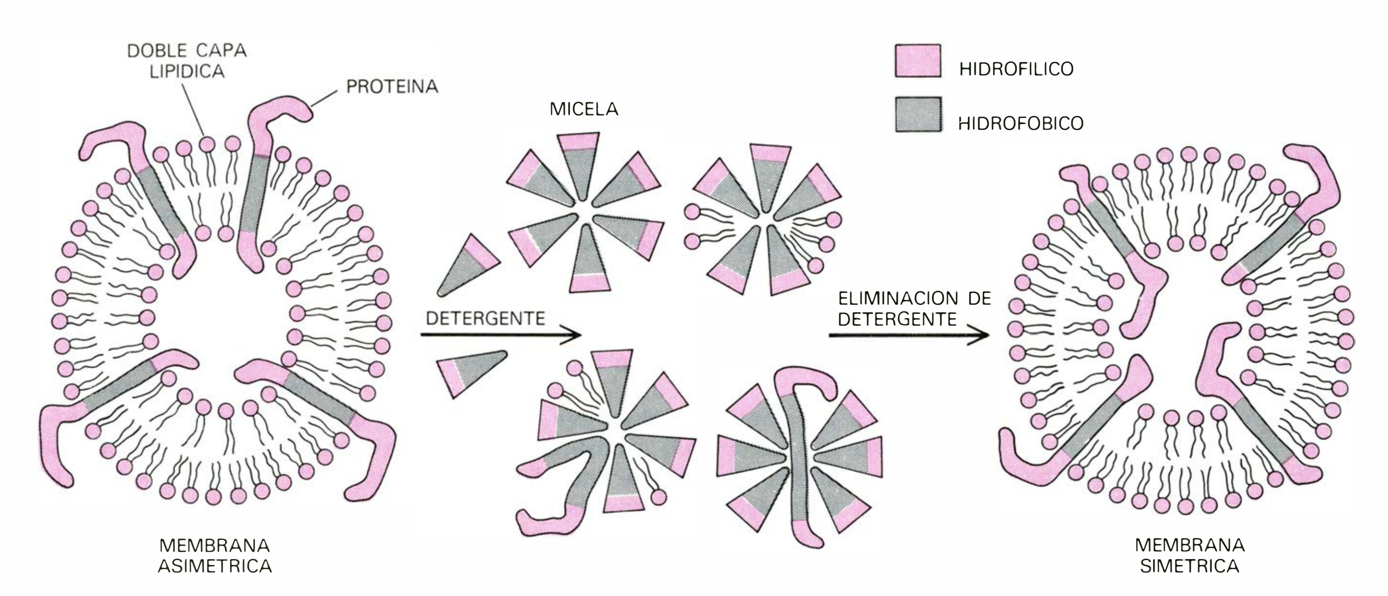
Sería una tarea formidable estudiar la síntesis de las proteínas de la membrana en una célula eucariota, e incluso en una procariota. La variedad de las proteínas es demasiado grande y demasiado reducido el número en que se presenta cada especie proteica. Nosotros y nuestros colegas elegimos un sistema de membrana muy simple, la que recubre externamente un virus; se trata de una membrana de fosfolípido tomada de la membrana plasmática de una célula animal infectada. El fragmento de membrana está densamente cubierto con proteínas constitutivas de la misma, pero todas ellas son de un mismo tipo, codificado por el genoma vírico (dotación completa de los genes víricos).

El virus estudiado es el de la estomatitis vesicular (VSV), que infecta a muchos animales de granja en los que provoca síntomas similares a los de los virus de la gripe. Cuando una partícula VSV penetra en una célula, detiene la síntesis normal de proteínas de la célula huésped y controla los enzimas y otros componentes del aparato sintético. En el período siguiente se forman sólo proteínas víricas. Al mismo tiempo, el genoma del virus, que es una cadena de ARN, se replica varias veces a costa de los enzimas codificados por el genoma del virus y contenidos en la partícula vírica. El ciclo se completa cuando los ARN replicados y las proteínas se unen en la membrana plasmática. Cada copia del ARN, junto con una dotación de proteínas, se envuelve en una porción de membrana y emerge de la célula para formar una nueva generación de viriones, o partículas víricas. En cuanto a composición lipídica, la membrana vírica tiene la misma que la de la célula huésped, pero la membrana incorpora sólo proteínas especificadas por el virus.

El genoma VSV codifica únicamente cinco polipéptidos, y todos ellos están en la partícula vírica madura. Tres polipéptidos están fuertemente asociados al ARN genómico; reciben el nombre de



MODELO DE MEMBRANA PLASMATICA, que encierra proteínas, hidratos de carbono y lípidos. Las proteínas constitutivas se insertan en la doble capa lipídica; las proteínas periféricas guardan mera asociación con la superficie de la membrana. Los carbohidratos están constituidos por monosacáridos, o azúcares sencillos, que se polimerizan en cadenas, que se ligan a las proteínas (para formar glucoproteínas) o a los lípidos (para formar glucolípidos). La asimetría de la membrana se manifiesta por diversos aspectos. Los carbohidratos se localizan siempre en la superficie exterior y las proteínas periféricas se encuentran casi siempre en la superficie citoplasmática, o interior. Las dos capas sencillas de lípidos comprenden diferentes proporciones de varios tipos de moléculas lipídicas. Y, lo que es más importante, cada especie de proteína constitutiva tiene una orientación definida, la misma para las moléculas pertenecientes a esa especie. Las proteínas son polímeros formados por aminoácidos unidos en secuencia lineal.



AUTOENSAMBLAJE DE UNA MEMBRANA. El autoensamblaje conserva la estructura básica de ésta, pero no su asimetría. Una membrana puede romperse por efecto de una alta concentración de detergente, que es una molécula anfipática que da lugar a pequeñas formas goticulares denominadas micelas. El detergente disuelve los componentes de la membrana envolviendo las porciones hidrofóbicas de lípidos y proteínas con las micelas, de modo que queden protegidos del contacto con el agua. Si el detergente se elimina luego, los lípidos desarrollan espontáneamente una nueva doble capa, incorporando en ella las proteínas constitutivas. Las proteínas asumen en general orientaciones aleatorias. Los experimentos realizados en este sentido han demostrado que las membranas de la célula no pueden autoensamblarse. Las proteínas constitutivas deben insertarse en una membrana que exista previamente y que tenga una lateralidad definida. Es decir, se requiere que la proteína se inserte siempre por la misma cara.

nucleoproteínas, por analogía con ciertas proteínas del núcleo celular de los eucariotas. Dos de las nucleoproteínas son enzimas responsables de la replicación y transcripción del genoma vírico; la tercera nucleoproteína, llamada *N*, tiene probablemente función estructural, y es la más abundante de las tres. Una cuarta proteína, designada por *M*, de matriz, se sintetiza como una proteína citoplasmática soluble, pero a continuación se incorpora al virión como una proteína periférica sobre la superficie interna de la membrana. La última proteína, llamada *G*, de glucoproteína, ha constituido el centro de nuestras investigaciones. *G* es una proteína de membrana que abarca el doble nivel lipídico; en micrografías electrónicas, se observa como una serie de espículas distribuidas sobre la superficie del virión. Puede detectarse también en la membrana plasmática de las células infectadas durante el proceso de extrusión. No se ha determinado la secuencia de aminoácidos y la estructura tridimensional de *G*, pero sabemos que comprende alrededor de 550 aminoácidos; hay incluidas, además, dos cadenas laterales de carbohidratos. La orientación de *G* en la doble capa de lípidos es indiscutiblemente asimétrica; la mayor parte de los polipéptidos y todos los azúcares se exponen en la cara externa de la membrana, y sólo un grupo de unos 30 aminoácidos sobresalen de la cara citoplasmática.

Las fases involucradas en la extrusión del virión no se conocen con certeza, pero puede proponerse una secuencia de fenómenos. El ARN genómico y las nucleoproteínas se unen evidentemente en el citoplasma, formando una estructura denominada nucleocápsida. Al mismo tiempo, *G* aparece en la membrana plasmática. Es la única glucoproteína de membrana que la célula sintetiza en grandes cantidades, por lo que algunas partes de la membrana probablemente adquieren una cubierta densa de dicha proteína. Se supone que la proteína matriz *M* sirve de puente de unión entre la nucleocápsida y el grupo de aminoácidos citoplasmáticos de *G*. Si esto es así, es fácil sospechar un mecanismo de extrusión. Siempre que una nucleocápsida se acerca a la membrana, las proteínas *M* forman uniones en el punto del primer contacto entre las nucleoproteínas y el grupo citoplasmático de *G*. Como resultado de la atracción entre estos componentes, gran parte de la membrana está envolviendo la nucleocápsida y está unida por la proteína matriz. Podría incorporarse cualquier proteína de membrana de la célula huésped que estuviera en la proximidades. Al final, la nucleocápsida entera está envuelta y el virión abandona la célula. No es que la nucleocápsida haya sido empujada a través de la membrana, sino más bien atraída hacia ella por las uniones de la proteína matriz.

Las proteínas víricas se forman por un proceso similar al de las proteínas de la célula huésped. El código que especifica la secuencia de aminoácidos de cada una de las cinco proteínas víricas lo porta un fragmento de ARN mensajero que se sintetiza como una copia de una región del ARN del genoma vírico. El ARN mensajero se traduce en un polipéptido a través de una estructura subcelular denominada ribosoma. El ribosoma está formado por varias docenas de proteínas y tres o más tipos de ARN organizados en una subunidad grande y otra pequeña. El ARN mensajero se desplaza a través del ribosoma y el polipéptido en crecimieno se ensambla en una subunidad grande. Cada aminoácido incorporado al polipéptido es transportado por una molécula de ARN de transferencia, que reconoce una secuencia particular de tres nucleótidos del ARN mensajero. La traducción comienza con el terminal amino y cursa hacia los terminales carboxilo de la proteína. El punto donde los aminoácidos se incorporan a la cadena en crecimiento está probablente incluido en la subunidad grande; durante toda la traducción, unos 40 aminoácidos, los cuarenta incorporados más recientemente, están ocultos en el interior del ribosoma. Conforme la cadena crece, estas unidades son sucesivamente empujadas fuera por la nuevas que se van incorporando; una vez completa la molécula polipeptídica, se libera la proteína entera. En general, una molécula de ARN mensajero se traduce simultáneamente en varios ribosomas, produciéndose en cada uno de ellos una cadena polipeptídica.

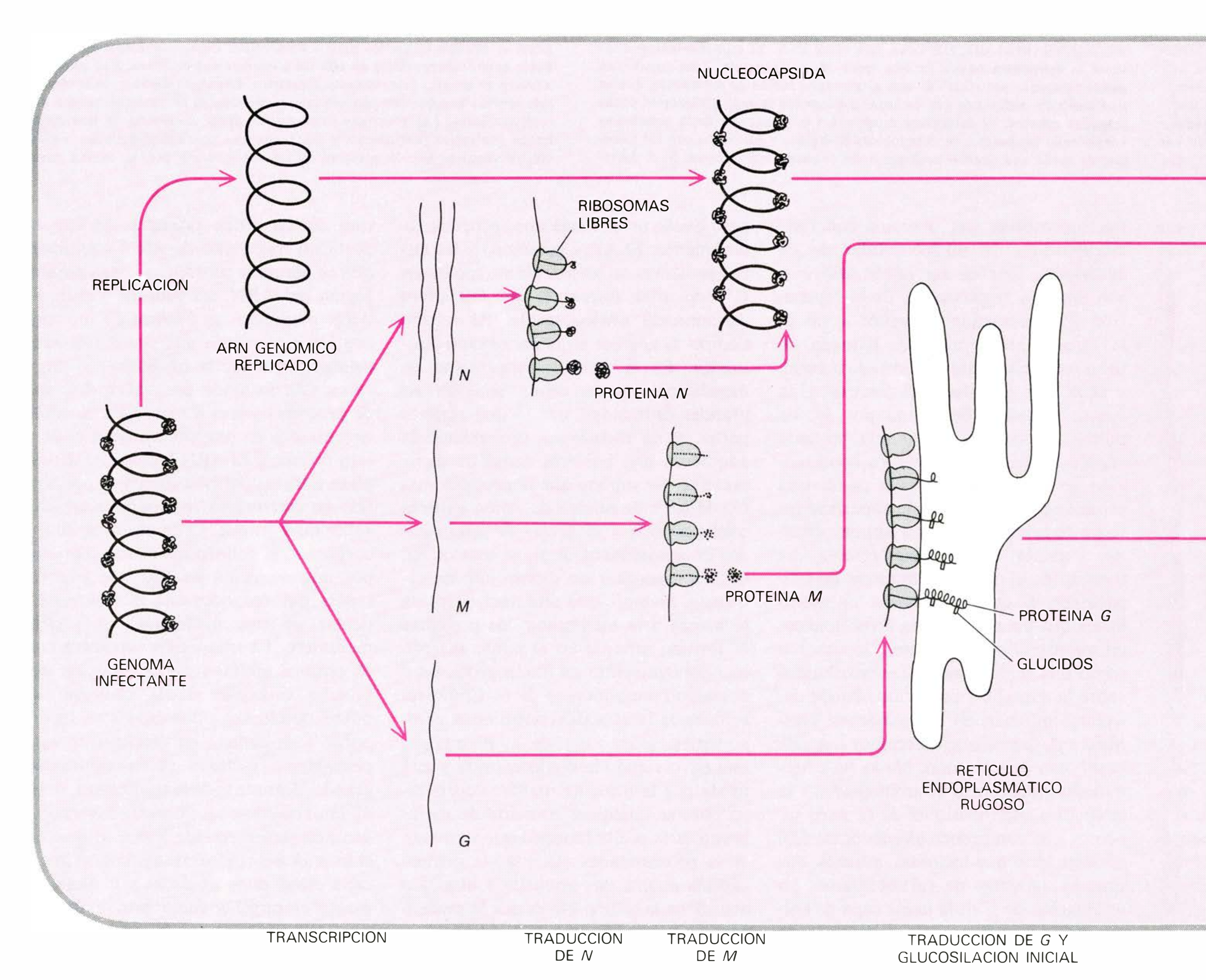
Se ha señalado con frecuencia la exis-

tencia, en las células eucariotas, de dos tipos de ribosomas; algunos están libres en el citoplasma, mientras que otros están ligados a las membranas del retículo endoplasmático. Hoy se sabe que los ribosomas de ambas clases tienen idéntica estructura y son intercambiables. Un ribosoma de los que están ligados a la membrana puede liberarse y desplazarse a otro punto de la membrana. Pero ambos tipos ribosómicos difieren en su función. Los ribosomas libres sintetizan principalmente proteínas solubles, mientras que los ligados a la membrana sintetizan proteínas constitutivas de la membrana y proteínas destinadas a la secreción celular.

La proteína matriz del VSV se forma en ribosomas libres. Ello no debe sorprendernos, ya que, inmediatamente después de la síntesis, la proteína se libera por sí misma en el citoplasma; se liga a la membrana (en la cara citoplasmática) sólo cuando se inicia la extrusión. También se ha demostrado que otras proteínas periféricas se originan como proteínas solubles, y, según parece, todas las proteínas ligadas a la superficie citoplasmática, aunque no incrustadas allí, se sintetizan de la misma forma.

La glucoproteína *G* sigue una trayectoria más complicada. Sintetizada por los ribosomas del retículo endoplasmático, se transporta posteriormente a la membrana plasmática, con una parada como mínimo a lo largo del proceso. El retículo endoplasmático constituye una red de membranas que se encuentran cerca del núcleo celular, a menudo rodeándolo parcialmente. Las membranas están dispuestas en capas; en las micrografías electrónicas, cada capa aparece como una vesícula alargada y plana, una célula dentro de la célula.

La glucoproteína del VSV surge primero en la célula como un componente del retículo endoplasmático rugoso, que debe el adjetivo rugoso a la gran cantidad de ribosomas adosados a su superficie. *G* se incrusta en la membrana de una vesícula del retículo, con la espícula grande, que lleva dos cadenas

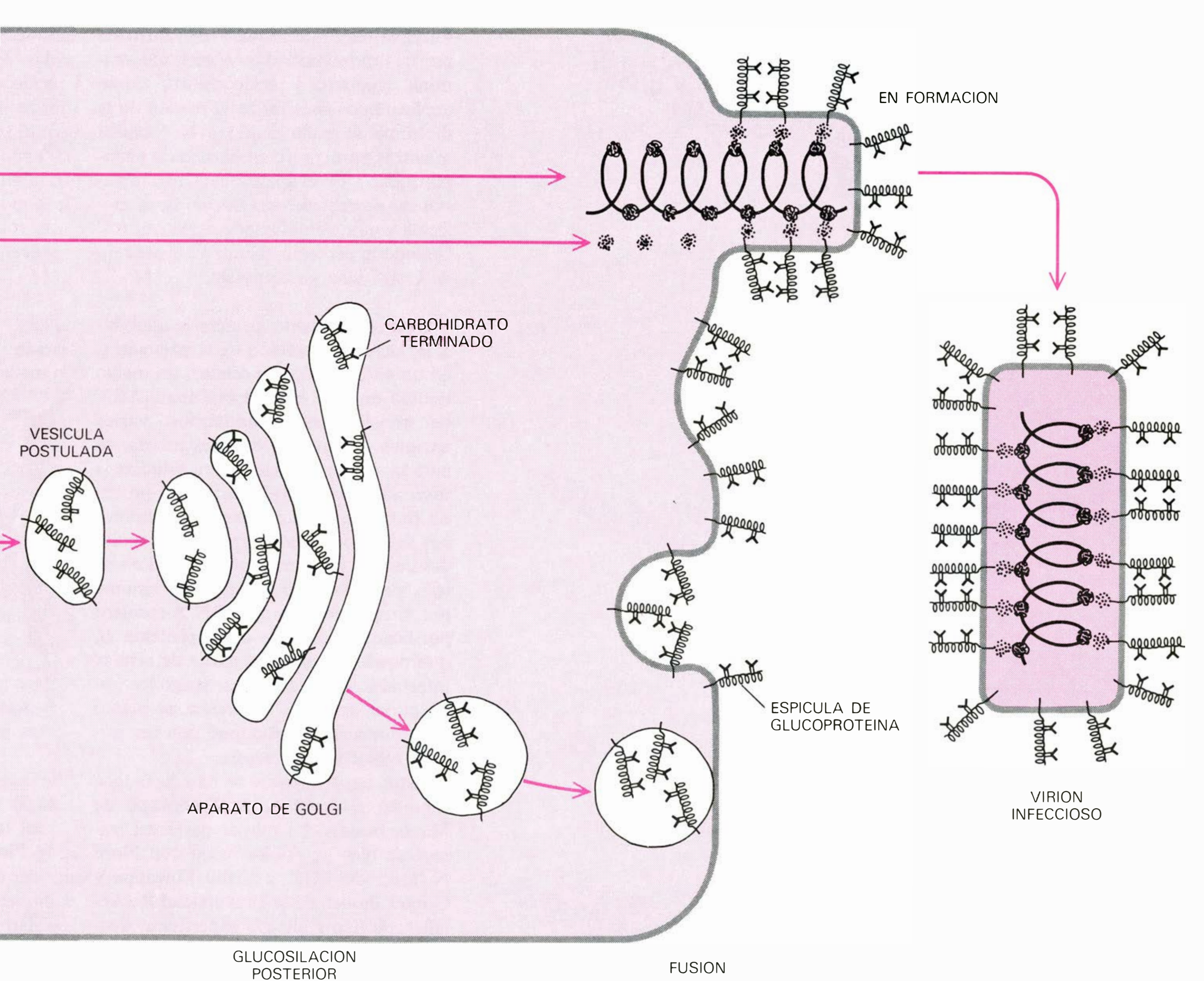


PROTEINAS VIRICAS. Se fabrican utilizando la maquinaria genética de la célula, pero su estructura viene especificada por el genoma del virus. El ciclo infeccioso del VSV empieza cuando un virus inserta su material genético, una cadena de ARN, en el interior celular. El ARN genómico se transcribe por dos enzimas víricos en cinco moléculas de ARN mensajero. Luego, el ARN genómico se replica por enzimas víricos similares. Cada cadena de ARN mensajero se traduce, a través de unos orgánulos denominados ribosomas, en muchas proteínas idénticas; al haber cinco moléculas de ARN mensajero, el genoma vírico codifica cinco proteínas. Tres son nucleoproteínas; entre éstas se cuentan los dos enzimas necesarios para la transcripción y replicación y una tercera proteína, denominada *N*, cuya principal función es quizá de tipo estructural. Las tres nucleoproteínas se enlazan con el ARN genómico para formar un complejo llamado nucleocápsida. Una cuarta proteína, llamada *M*, de matriz, es inicialmente un componente soluble del citoplasma. Las cuatro proteínas se sintetizan en los ribosomas que están libres en el citoplasma. La última proteína, denomi-

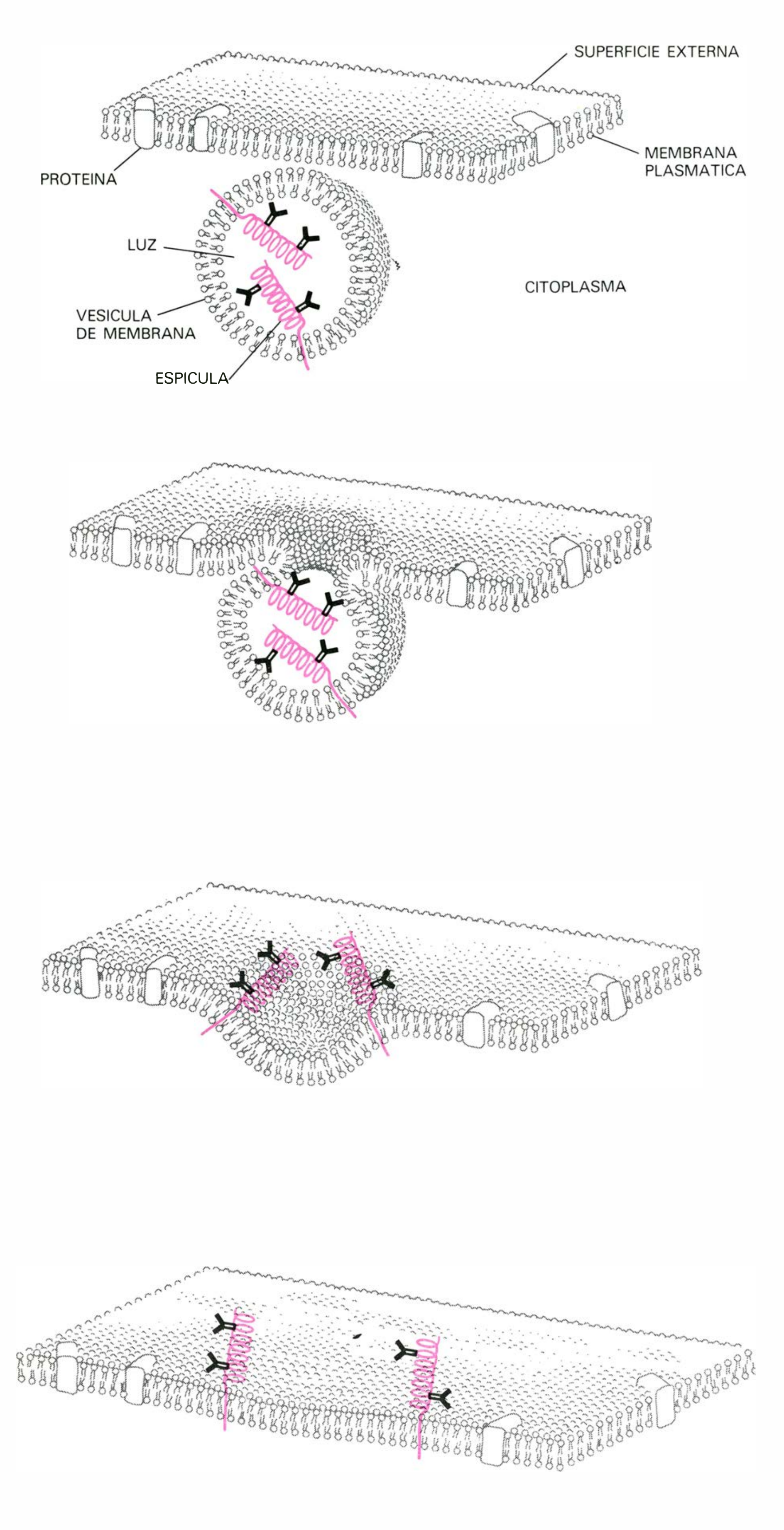
de glúcidos, sobresaliendo en la luz, o cavidad interna, de la vesícula. La parte corta del terminal carboxílico de la proteína mira hacia el citoplasma. Transcurridos 20 o 30 minutos, *G* puede localizarse en el aparato de Golgi, que es otro orgánulo formado por muchas vesículas cerradas membranosas. Al aparato de Golgi se le llama a menudo orgánulo membranoso liso, pues carece de ribosomas adosados. Sólo después de que la proteína haya pasado a través del aparato de Golgi alcanzará la membrana plasmática.

Se desconoce cuál es el medio de transporte entre estos orgánulos. Quizá lo más probable sea que las pequeñas vesículas empiezan a desarrollarse a partir de una membrana cerrada y posteriormente se funden con otra membrana, llevando la proteína consigo y conservando siempre su orientación. Desde este punto de vista, importa no olvidar la orientación de la proteína a lo largo de todo el trayecto. En principio parece como si la orientación de la proteína se invirtiera en algún punto del proceso. En el retículo endoplasmático la proteína se dirige hacia el interior, hacia la luz de la vesícula, pero en la membrana plasmática lo hace hacia el exterior de la célula. Pero, en realidad, esto no representa ninguna inversión, sino el mantenimiento de la misma orientación. Adviértase que la parte de proteína que hemos llamado "el pequeño grupo" está siempre expuesta en el citoplasma, y la otra porción (la espícula) siempre está fuera del citoplasma. El lumen de un orgánulo, como el del retículo endoplasmático o el aparato de Golgi, es topológicamente equivalente al exterior de la célula, hecho que George E. Palade, de la Facultad de Medicina de la Universidad de Yale, demostró por vez primera. En una perspectiva topológica, el extremo espicular de la proteína se halla ya fuera de la célula desde el momento en que penetra en el retículo endoplasmático.

Las cadenas laterales glucídicas de la glucoproteína se fijan antes de que se alcance la membrana plasmática. En el virión, la proteína lleva dos estructuras



nada *G*, por ser una glucoproteína, posee cadenas laterales de carbohidratos (cuya estructura no viene especificada por el genoma del virus). *G* se forma en los ribosomas asociados a membranas del retículo endoplasmático rugoso y, desde el momento en que aparece la primera proteína, está ligado a estas membranas. Los monosacáridos nucleares se añaden a la molécula de *G* en crecimiento en el retículo endoplasmático rugoso, y, después de 20 o 30 minutos, las cadenas de carbohidratos se completan en el aparato de Golgi. Aunque se desconoce el mecanismo de transporte entre estos orgánulos, puede relacionarse con el de vesículas que abandonan una membrana para fundirse con otra. La glucoproteína podría alcanzar la membrana plasmática, en donde de manera similar forma espículas sobre la superficie de la célula. El ciclo vital del virus se completa cuando la nucleocápsida se cubre con una porción de la membrana plasmática que posee proteína *G*; esta partícula vírica recientemente ensamblada emerge de la célula. La salida está mediatizada por la proteína de la matriz; ésta podría mantener unidas las nucleoproteínas y un apéndice de *G* que se asoma al citoplasma.



FUSION DE UNA VESICULA con la membrana plasmática. Dicha fusión conserva la orientación de cualquier proteína constitutiva inscrita en la doble capa de la vesícula. Inicialmente, las grandes espículas terminales de la proteína *G* miran hacia la luz, o cavidad interna de las vesículas. Después de la fusión, la espícula está sobre la superficie externa de la membrana plasmática. La orientación no ha variado, por cuanto el otro extremo de la molécula, el grupo pequeño de aminoácidos está inmerso en el citoplasma. El lumen de la vesícula y el exterior de la célula son topológicamente equivalentes.

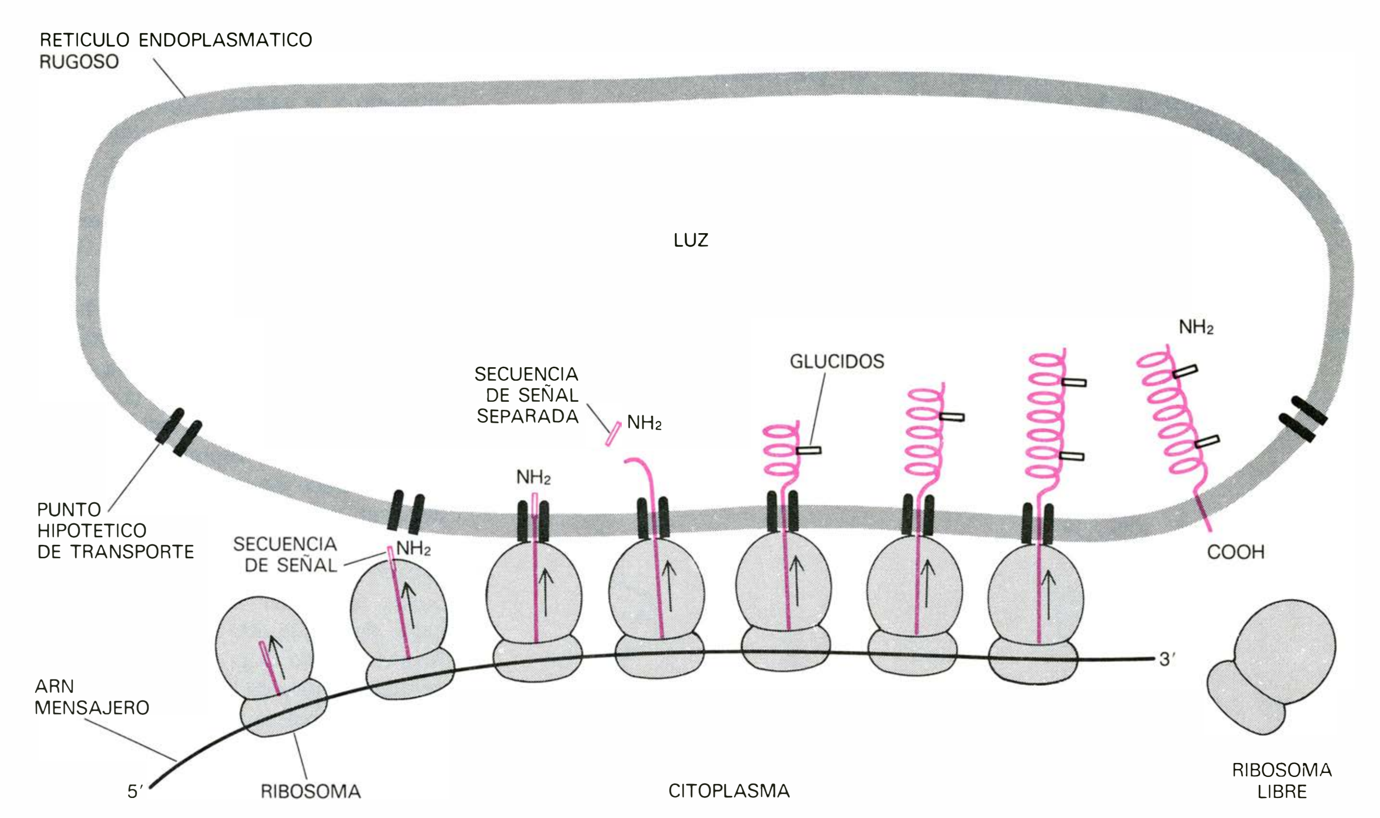
carbohidratadas idénticas, cada una formada por una docena, más o menos, de monosacáridos, o unidades de azúcar ligados. Se ignora la función de las cadenas de azúcares en la glucoproteína del VSV, pero su estructura recuerda a muchas glucoproteínas normales presentes en el suero sanguíneo.

Donald F. Summers y sus colaboradores, de la Universidad de Utah, han determinado la secuencia de monosacáridos de las cadenas de carbohidratos y han demostrado que cada cadena puede considerarse como formada por dos partes. La región inmediata a la proteína, y unida directamente a ella, llamada núcleo, consta exclusivamente de los monosacáridos manosa y *N*-acetilglucosamina. La parte glucídica alejada constituye la región terminal y está formada por tres monosacáridos: *N*-acetilglucosamina, galactosa y ácido siálico. Como explicaremos más tarde, el núcleo de la molécula se halla unido a la proteína mientras permanece en el retículo endoplasmático. En el aparato de Golgi, algunos monosacáridos del núcleo de la molécula son sustituidos por otros nuevos. Cuando la proteína abandona el aparato de Golgi, está ya completa.

Hemos estudiado la síntesis, glucosilación e inserción de la proteína *G* en un sistema libre de células, un medio líquido en que había ribosomas, ARN de transferencia, aminoácidos, varios enzimas y otros ingredientes necesarios para la síntesis proteica. Para estudiar la inserción y la glucosilación de la proteína hubo que añadir, además, membranas del retículo endoplasmático despojadas de sus ribosomas endógenos. La síntesis proteica se inicia en este sistema por simple adición de ARN mensajero purificado, específico de la proteína *G*, que puede obtenerse a partir de células infectadas. Para identificar luego los polipéptidos de nueva formación, se marcó un aminoácido (metionina) con un isótopo radiactivo del azufre.

Estos experimentos se han llevado a cabo en el Instituto de Tecnología de Massachusetts. La mayor parte del trabajo se hizo en colaboración con Flora N. Katz, del MIT, y Vishu Lingappa y Gunter Blobel, de la Universidad Rockefeller. Se han realizado experimentos similares en la Universidad de McMaster por parte de Hara Ghosh y Frances Toneguzzo. Presentaremos primero algunas conclusiones derivadas de nuestros hallazgos y discutiremos luego los experimentos en sí mismos.

La tesis central de nuestro modelo es que las proteínas constitutivas de la membrana se insertan en la doble capa



SINTESIS, GLUCOSILACION E INSERCION de la proteína *G*. **Se trata de procesos que están estrechamente relacionados. La proteína, o polipéptido, está formada por aminoácidos unidos en una secuencia lineal, especificada por el ARN mensajero. La síntesis comienza con el terminal amino (NH_2) del polipéptido. Entre los primeros 30 aminoácidos de *G* está la secuencia de señal que identifica la proteína como aquella que debe insertarse en la membrana del retículo endoplasmático rugoso. Debido a que unos 40 aminoácidos permanecen ocultos en el ribosoma, la secuencia de señal no emerge hasta que el polipéptido tiene un tamaño del orden de 70 aminoácidos. Al mismo tiempo, la secuencia de señal es reconocida por alguna molécula, presumiblemente una proteína, en la membrana del retículo endoplasmático. Se cree que esta proteína hipotética facilita el paso del polipéptido a través de la doble capa lipídica. Una vez en el lumen del retículo, la secuencia de señal se separa. La proteína continúa alargándose (creciendo); conforme va desarrollándose, empuja a través de la membrana y se pliega en el lumen. Al ir penetrando, fija dos cadenas glucídicas idénticas, previamente formadas. Las proteínas segregadas por la célula pasan todas así a través de la membrana; por razones que no se comprenden perfectamente, *G* se queda inmóvil durante el tiempo en que completa la traducción, con unos 30 aminoácidos remanentes en el citoplasma. De este modo, la glucoproteína completa tiene su terminal amino, la mayor parte de su volumen y todos los carbohidratos en la luz del retículo, y un reducido grupo de aminoácidos que incluyen su terminal carboxílico (COOH) en la cara citoplasmática. Una vez la proteína ha adquirido su configuración definitiva en plegamiento, no puede salir fuera de la membrana, ni puede efectuar ningún movimiento de "flip-flop" (sistema biestable) se halla anclada en una orientación asimétrica. Los componentes de la célula no se muestran a escala: el ribosoma es unas cincuenta veces mayor que la proteína *G*.**

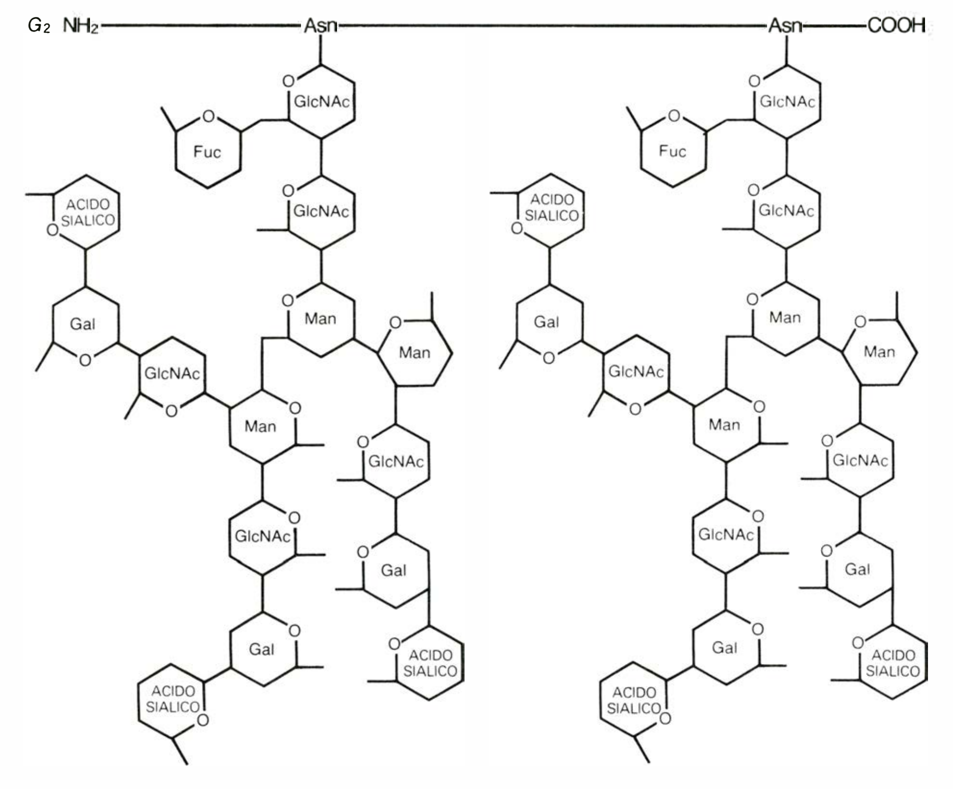
lipídica a medida que se van sintetizando, y antes de que adopten su configuración definitiva, en plegamiento. Este modelo permite comprender dos cuestiones importantes. En primer lugar, ayuda a explicar cómo la proteína puede cruzar la región hidrofóbica de la membrana: pasando a su través antes de que el polipéptido plegado haga más difícil la inserción. En segundo lugar el modelo explica la orientación asimétrica de las moléculas proteicas: la proteína puede estar insertada en una única dirección (la dirección de la elongación del polipéptido) y sólo a partir de una cara de la membrana (la cara del citoplasma con los ribosomas unidos).

La traducción de *G* se parece a la de todas las proteínas: empieza con la unión del ARN mensajero a la subunidad menor del ribosoma; luego, se añade la subunidad mayor. En esta etapa, el ribosoma no está asociado con una membrana, sino que se encuentra libre en el citoplasma. Deben unirse alrededor de 40 aminoácidos antes de que el polipéptido empiece a emerger de la subunidad mayor del ribosoma; se trata de las 40 unidades de aminoácidos terminales de la molécula. En cuanto se añaden otros 30 aminoácidos a la cadena en crecimiento, sobresalen los 30 primeros. Esta porción del polipéptido forma un grupo de aminoácidos llamados secuencia de señal, que es esencial para la distinción entre proteínas de membrana y proteínas de secreción en una cara, y proteínas solubles en la otra. La secuencia de señal es reconocida por algún componente de la membrana del retículo endoplasmático, presumiblemente una proteína de membrana. Esta membrana transporta loci de unión al polipéptido naciente y contribuye a iniciar su paso a través de la doble capa.

La existencia de la secuencia de señal fue propuesta, por vez primera, por Cesar Milstein y George G. Brownlee y sus colegas, del Medical Research Council Laboratory of Molecular Biology de Cambridge (Inglaterra), basándose en estudios propios sobre la síntesis de anticuerpos proteicos. En favor de la secuencia, aportaron nuevas pruebas los experimentos de Blobel y colaboradores, quienes examinaron sobre todo proteínas destinadas a ser excretadas por la célula. Según parece, los mecanismos celulares que gobiernan los dos tipos de proteínas son iguales en muchos aspectos. Ambos son extruidos al interior del retículo endoplasmático; hallándose la diferencia principal en que todas las proteínas de secreción sufren el mismo proceso y son liberadas en la luz de la vesícula, mientras que las proteínas de la membrana permanecen incrustadas en la doble capa de fosfolípidos, antes de que acabe de formarse el polipéptido. Cuando una vesícula lleva tales proteínas incluidas en la membrana plasmática, las proteínas de secreción se liberan fuera de la célula, mientras que las proteínas constitutivas permanecen unidas a la membrana. En proteínas de secreción, una vez la secuencia de señal ha dirigido el polipéptido a través de la doble capa lipí-

dica, se separa por acción de los enzimas del retículo; recientemente, se ha encontrado que la secuencia de señal de la proteína *G*, formada por 16 aminoácidos, se elimina también. Las secuencias de señal de varias proteínas de secreción y de *G* son distintas, pero comparten ciertas propiedades; en particular, admiten un predominio de aminoácidos hidrofóbicos.

Sólo cuando la secuencia de señal arrastra el polipéptido al retículo endoplasmático, el ribosoma se asocia a una membrana. Al principio, el ribosoma no está quizá ligado directamente a la membrana, sino que tan sólo se encuentra sujetado a la cadena polipeptídica. Más tarde, puede establecerse una interacción electrostática débil entre el ribosoma y la membrana.



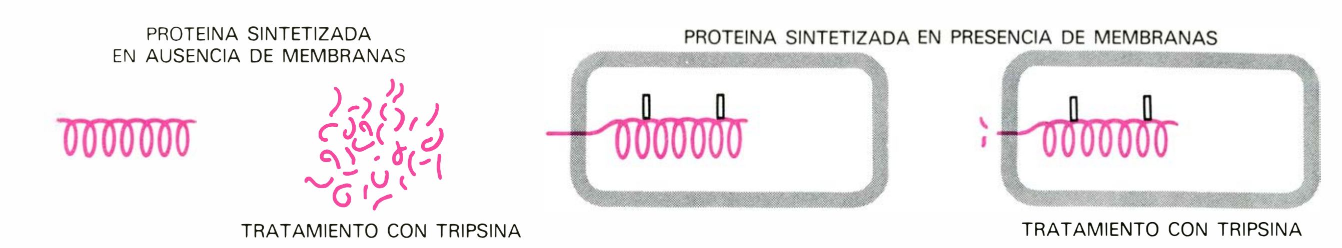
CADENAS LATERALES DE CARBOHIDRATOS. **Se construyen sobre la proteína *G* en dos pasos. En el retículo endoplasmático rugoso, la proteína en crecimiento incorpora las regiones centrales de dos cadenas de azúcares, que están formadas sólo por los monosacáridos manosa (Man) y *N*-acetilglucosamina (GLcNA o NAG), en el aparato de Golgi se modifican las regiones centrales (aunque no se eliminan) y se unen los carbohidratos terminales; éstos incluyen más *N*-acetilglucosamina, así como los monosacáridos galactosa (Gal) y ácido siálico. También se incorpora fucosa (Fuc). La glucoproteína completa tiene dos cadenas de carbohidratos idénticas, cada una ligada al polipéptido a través de una derivación en el aminoácido asparagina (Asn). Se cree que las dos unidades de asparagina se hallan situadas alrededor de las posiciones 150 y 400 a partir del terminal amino del polipéptido. Hay pruebas de que los carbohidratos centrales se incorporan a la proteína como una unidad, habiéndose ensamblado previamente sobre un glucolípido (molécula lipídica que puede contener uno o más residuos glucídicos).**

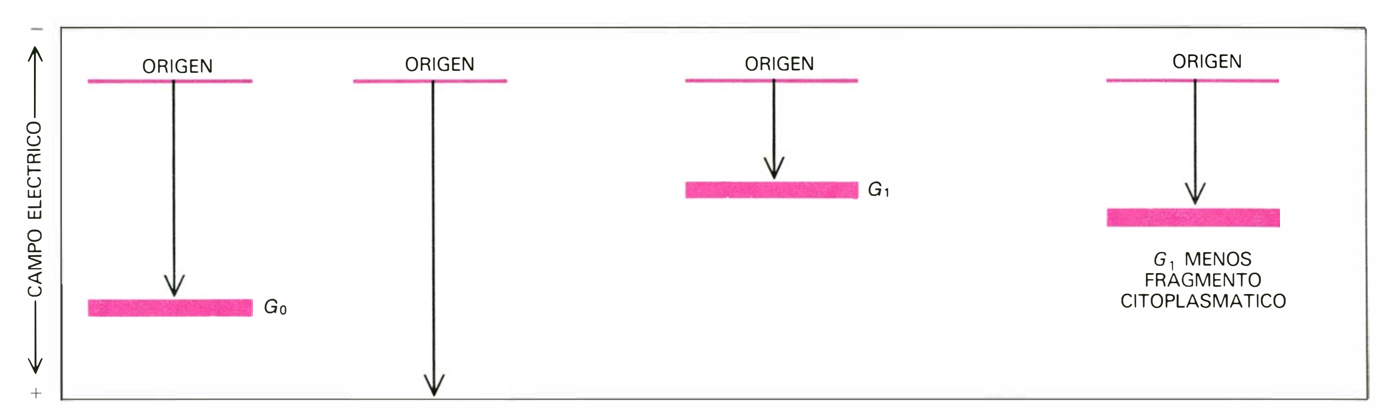
Con toda posibilidad, la proteína no penetra la doble capa lipídica directamente, sino que es ayudada por alguna otra proteína constitutiva que se halle ya en la membrana, y que se difunde hacia fuera cuando la inserción se ha completado. A medida que la cadena alcanza la luz del retículo, el polipéptido se configura, plegándose espontáneamente, tal como lo hace una proteína soluble para alcanzar una configuración de mínima energía en solución acuosa. Una vez ha llegado a este estado, la proteína no puede retroceder a través de la doble capa, dado que está fijada como un remache. Queda por explicar por qué el pequeño grupo de aminoácidos del terminal carboxílico del polipéptido no se desplaza por la cara interna, como sucede en las proteínas de secreción. Tampoco se ha establecido qué cantidad de energía adicional se requiere para transportar la proteína incipiente a través de la membrana.

En un sistema libre de células, donde las únicas membranas presentes son las del retículo endoplasmático, no se puede construir una glucoproteína completa, porque la modificación de la porción de carbohidratos sólo se ultima en el aparato de Golgi. La proteína *G* sintetizada artificialmente adquiere las dos cadenas de azúcares centrales, que parecen ser idénticas a las formadas en la célula intacta.

Cada una de las dos estructuras carbohidratadas está unida a un grupo lateral del aminoácido asparagina. Se supone que una unidad de asparagina está a unos 150 aminoácidos del terminal amino y la otra está a unos 400 del mismo extremo. Los carbohidratos centrales se ligan probablemente a las unidades de asparagina al poco de emerger por la cara del lumen. No se sintetizan sobre la proteína azúcar por azúcar, sino que se añaden todos de una vez. Las regiones centrales se forman de antemano sobre un lípido de membrana y se transfieren a *G* como una unidad. Puesto que las membranas reticulares empleadas en estos experimentos proceden de células que no habían sido infectadas por VSV, las cadenas glucídicas deben ser componentes de la célula normal, y no estructuras codificadas por el virus.

La glucoproteína creada por esta serie de procesos tiene sus terminaciones amino y gran parte de su masa, que en-





SISTEMA LIBRE DE CELULAS para la síntesis de *G*. Se empleó para someter a prueba la hipótesis de que la proteína se inserta en la membrana del retículo a medida que se desarrolla el polipéptido. El sistema es un medio líquido que contiene ribosomas, aminoácidos, varios enzimas y otras moléculas necesarias para el crecimiento proteico. La síntesis se inicia con la adición, al medio, de ARN mensajero que codifica para *G*. En este sistema, la proteína puede crecer bajo diferentes condiciones y, los resultados, analizados por electroforesis en gel, que separa las moléculas en razón de su tamaño; las moléculas pequeñas se desplazan en el gel más lejos que las mayores. Si *G* se forma en ausencia de membranas, el producto es una proteína llamada G_0, que sería normal si no fuera porque carece de carbohidratos y retiene la secuencia de señal del terminal amino. Repitiendo el procedimiento y añadiendo el enzima digestivo tripsina una vez completada la síntesis proteica, se reduce la proteína a multitud de pequeños fragmentos, que se desplazan fuera del gel. Cuando se añaden mambranas del retículo endoplasmático al medio, antes de que comience la síntesis proteica, se obtiene un producto distinto: la proteína G_1, que tiene la región central de ambas cadenas de carbohidratos y está desprovista de secuencia de señal. Cuando se añade la tripsina en este proceso sólo se degrada el grupo citoplasmático, porque la tripsina no puede atravesar la doble capa de lípidos. El resultado de todas estas experiencias demuestra que *G* adquiere las cadenas glucídicas sólo si entra en el retículo endoplasmático.

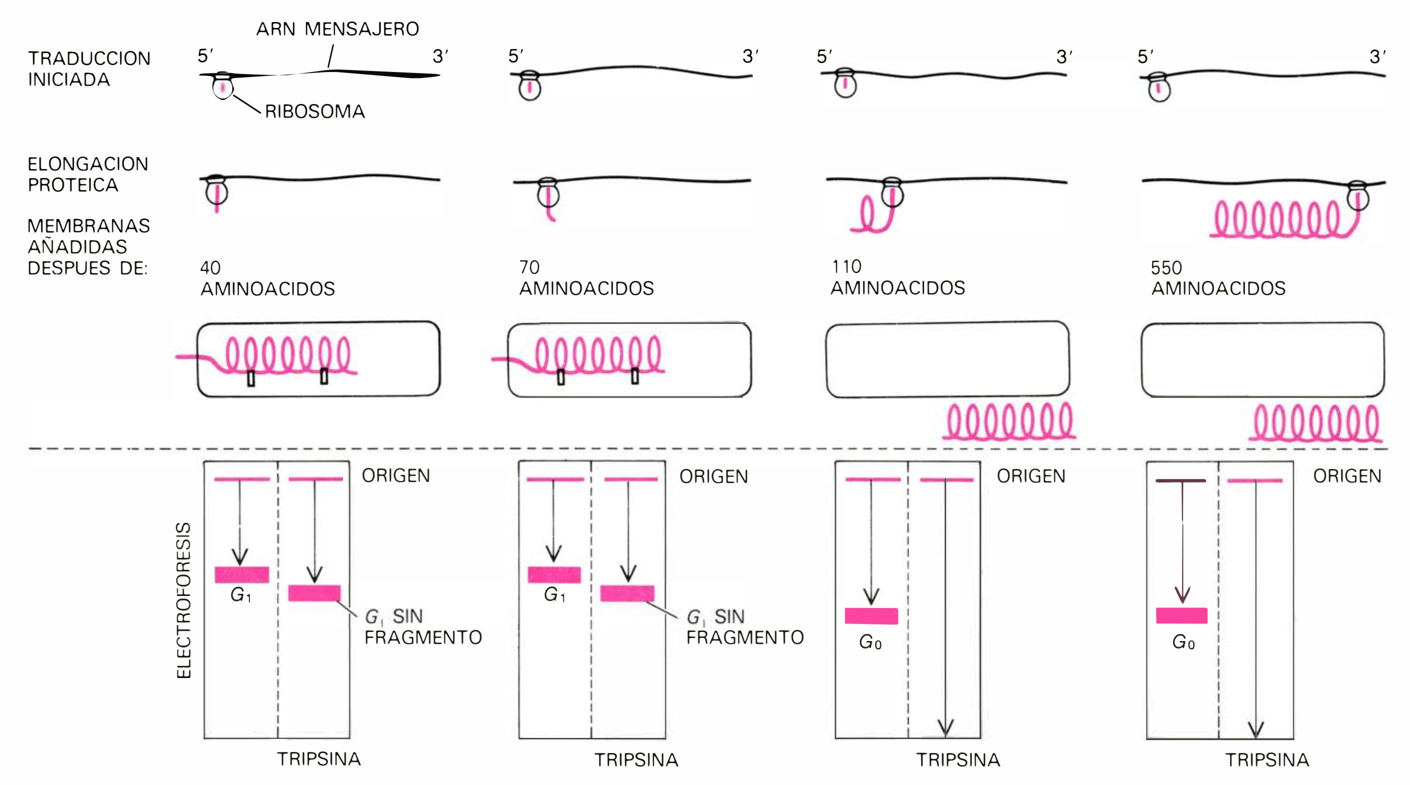
globa unos 500 aminoácidos, en la cara del lumen de la membrana reticular. Ambas cadenas glucídicas se hallan sobre la misma cara; es este extremo de la molécula el que eventualmente se observa como una espícula sobre la superficie exterior de la membrana plasmática y sobre el virión. Unos pocos aminoácidos, cuando menos, deben permanecer en la doble capa; y, aunque su número no se conoce con exactitud, se supone que oscilan entre 20 y 30. Sólo un grupo formado por las últimas 30 unidades, aproximadamente, incluido el terminal carboxílico, sobresale en la cara citoplasmática de la membrana.

¿Cómo se construyó y comprobó este modelo de la inserción de la proteína? En casi todos nuestros experimentos nos servimos de la técnica de la electroforesis en gel, que separa las proteínas en razón de su tamaño. Un experimento típico da una mezcla de proteínas, algunas en su forma original, otras reducidas a fragmentos por digestión enzimática y otras con sólo un complemento parcial de carbohidrato. La mezcla se coloca sobre un gel poroso; un campo eléctrico desplaza las proteínas a través del gel. Las moléculas pequeñas emigran a través de los poros del gel con mayor rapidez que las grandes; se separan así las distintas fracciones y pueden identificarse.

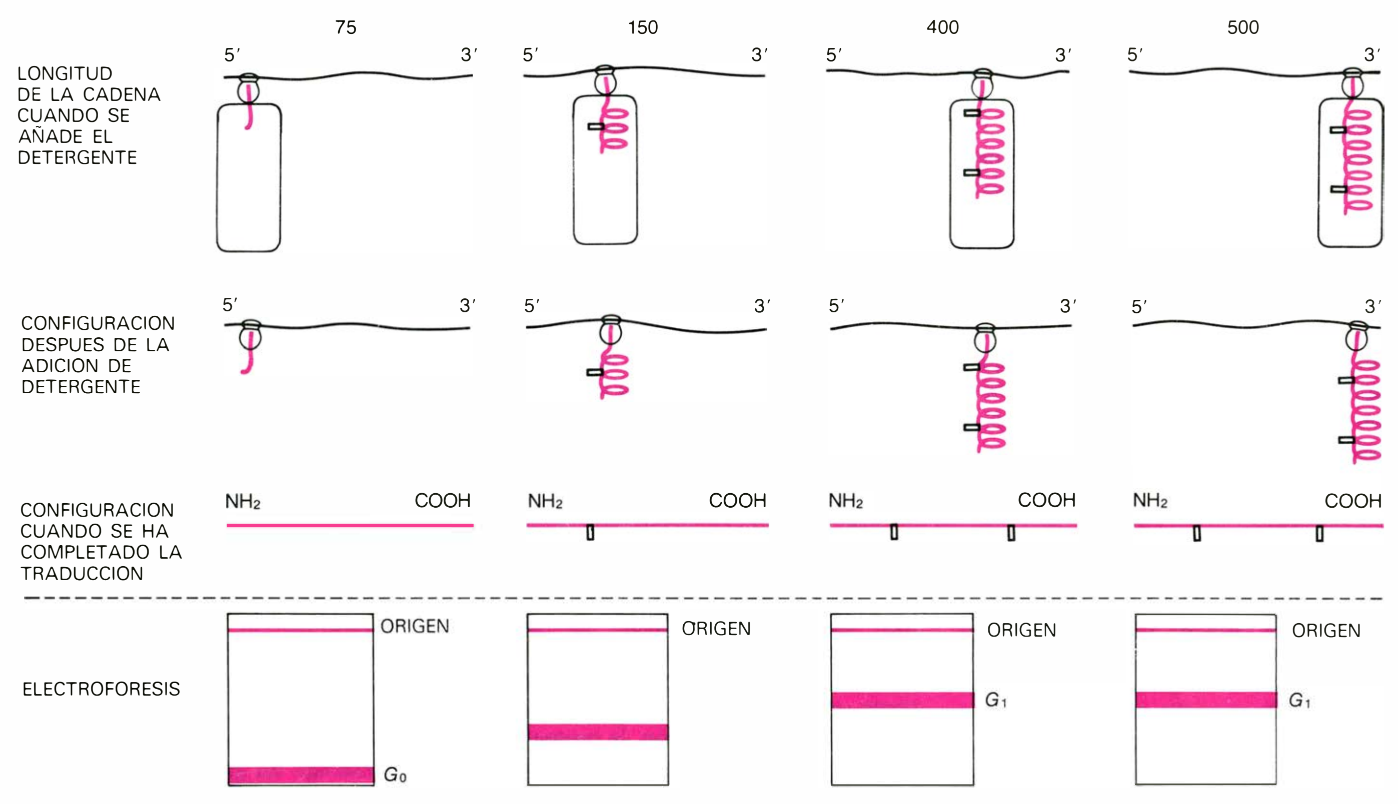
En un experimento se suministraron todos los enzimas y las restantes moléculas precisas para la síntesis normal de *G*, a excepción de membranas del retículo. Bajo tales condiciones, puede formarse el polipéptido en el ribosoma, pero evidentemente no puede insertarse en la membrana. La proteína resultante parece ser normal en todos los aspectos, menos en dos: carece de carbohidratos, ya que éstos pueden incorporarse sólo por acción de los enzimas del retículo, y mantiene la secuencia de señal de 16 aminoácidos en el terminal amino. Esa proteína desnuda recibe el nombre de G_0. Una proteína con sólo los carbohidratos centrales (y carente de secuencia de señal) se llama G_1, y G_2, la proteína terminada con ambos carbohidratos terminal y central. Aunque G_1 y G_2 sean intermediarios normales que se encuentran en células infectadas, G_0 no se forma durante la infección.

Si se añaden las membranas del retículo endoplasmático al medio antes de que empiece la traducción, la proteína se fabrica en la forma G_1, con ambas cadenas centrales de carbohidrato, pero sin secuencia de señal. Sin embargo se obtuvo un resultado eficaz cuando las membranas se añadían durante la síntesis. El modelo presentado antes predice que, a menos de que se añadan las membranas muy tempranamente, en la síntesis, la proteína formada no se insertará en la membrana. Parte del polipéptido debe ya haberse plegado en el citoplasma, y no podrá interaccionar con receptores de membrana ni atravesar la barrera de permeabilidad de la doble capa lipídica. Si añadíamos membranas una vez comenzada la síntesis, no ocurría nada en realidad; cuando el sistema se completó después, la proteína se mantuvo en la forma G_0.

Para comprobar la orientación de las moléculas, se sintetizó la proteína *G* en presencia de membranas; completada la síntesis proteica, se añadió tripsina a la solución. La tripsina es un enzima digestivo que puede romper un polipéptido por diversos puntos; ahora bien, dado que una proteína soluble no puede atravesar la membrana lipídica, sólo podrá digerir aquellas porciones de la molécula proteica que estén en la cara citoplasmática de la membrana. Todas las moléculas de *G* sometidas a este proceso presentan un gran fragmento de *G*, de unos 30 aminoácidos menos que el normal de G_1. Pudimos demostrar que únicamente se digería el terminal carboxílico del polipéptido, y que el gran fragmento protegido de la digestión enzimática llevaba las dos cadenas de azúcares.



ADICION DE MEMBRANAS una vez que la síntesis proteica empezó. Tal ensayo demostró que la inserción tenía que ser simultánea con la síntesis para que resultara con éxito. Si se deja que el polipéptido llegue a alcanzar 70 aminoácidos de longitud antes de que se introduzcan las membranas del retículo, la proteína puede penetrar en la membrana y, en consecuencia, desarrollarse normalmente. La electroforesis muestra que adquiere los dos carbohidratos centrales y que sólo el pequeño fragmento citoplasmático se degrada por efecto de la tripsina. Con un retraso suficiente entre el inicio de la síntesis y la introducción de membranas, el plegamiento de la proteína sintetizada ha progresado demasiado para entrar en la doble capa. Debido a que el plegamiento expone al exterior muchos aminoácidos hidrofílicos, el polipéptido queda excluido permanentemente del retículo. Cuando se analizan las proteínas por electroforesis, se encuentra que carecen de carbohidrato y han sido digeridos completamente por el enzima tripsina.



DESTRUCCION DE LAS MEMBRANAS una vez comenzó la síntesis proteica. Con este ensayo se probó que los carbohidratos centrales pueden unirse al polipéptido sólo dentro de las membranas del retículo endoplasmático. La traducción de varias copias de *G* comenzó simultáneamente en un medio con abundantes membranas del retículo; se les permitió continuar durante períodos diversos antes de incorporar un detergente en el medio, que destruyera las membranas. Se dejó luego que la elongación del polipéptido finalizara en ausencia de membranas. Al romperse las membranas antes de que la cadena alcanzara una longitud de 150 aminoácidos, la única proteína que se halló fue G_0, que carece de todos los carbohidratos. Si no se añade detergente hasta que se hayan incorporado unos 400 aminoácidos o más, se produce G_1 normal con dos cadenas glucídicas. Aplicando el detergente a intervalos intermedios se producen proteínas que nunca se encuentran en la naturaleza, con sólo la primera cadena de carbohidratos.

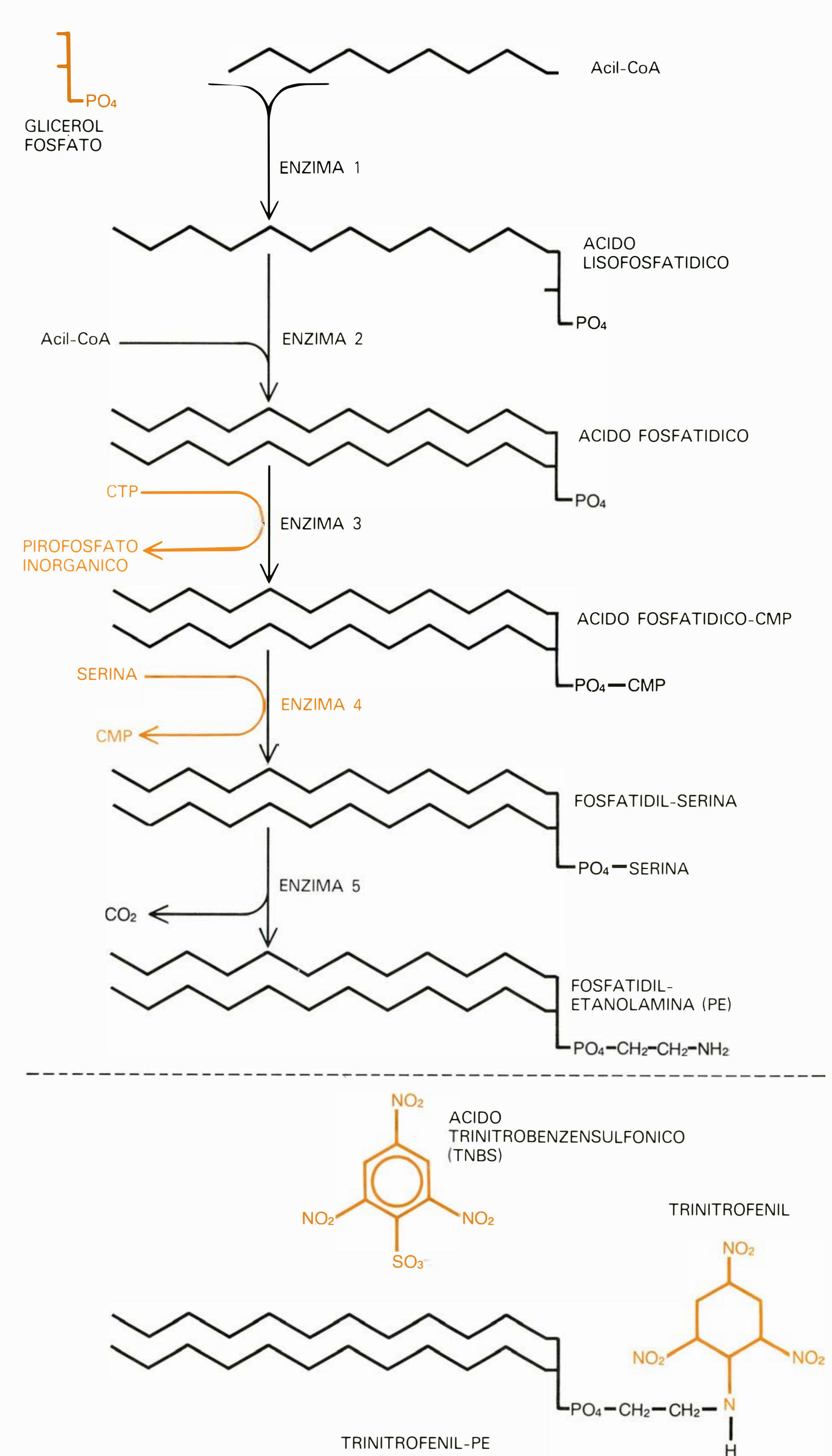
Estos hallazgos pusieron de manifiesto que sólo se proyectan hacia el citoplasma unas 30 unidades correspondientes al terminal carboxílico, y que el carbohidrato está en el lumen de la vesícula. Es más, el que todas las copias de *G* pierdan el mismo fragmento indicaba que todas las moléculas de *G* en la membrana presentaban la misma orientación.

Para esclarecer cuándo debía empezar exactamente la inserción en la doble capa, realizamos una serie de experimentos en los que progresivamente se sintetizaban segmentos de proteína cada vez más largos, antes de añadir las membranas al medio. Encontramos que, si se añadían las membranas antes de que estuvieran polimerizados los primeros 70 aminoácidos, el polipéptido se insertaba de forma correcta e iba a formar una G_1 normal. Si las membranas se añadían más tarde, sólo se formaban G_0 sin azúcares, y este polipéptido no se incluía debidamente en la membrana, sino que quizá se abandonaba en el citoplasma. Ya que unos 40 aminoácidos de los primeros 70 se hallan ocultos en el ribosoma, los primeros 30 deberán llevar la secuencia de señal.

Un experimento que fue, en algunos aspectos, el anverso del anterior probó que se va impulsando la proteína a través de la membrana a medida que cursa su proceso de elongación. La interpretación del experimento se basaba en la hipótesis de trabajo de que el carbohidrato puede añadirse sólo en la luz del retículo endoplasmático.

Se dejó crecer en distintos segmentos de polipéptido en presencia de membrana; se destruyó la membrana disolviéndolos con un detergente. El detergente no incide en la síntesis proteíca, pero sí impide la unión de los glúcidos, debido a que los enzimas que transfieren las cadenas glucídicas laterales se ligan a la cara, de la membrana del retículo endoplasmático, que da al lumen y se separan de los ribosomas a raíz del tratamiento con el detergente. Encontramos que cuando se añadía el detergente, antes de que el polipéptido tuviera una longitud del orden de 150 aminoácidos, y se completara la proteína, sólo se observaba proteína G_0. Así pues, los carbohidratos no se incorporan antes de que el polipéptido tenga 150 aminoácidos de longitud. Si la membrana se mantenía intacta hasta que se hubiera incorporado unos 400 aminoácidos a la cadena, la glucoproteína se completaba en la forma G_1 con los dos carbohidratos centrales.

Esto significaba que todos los azúcares centrales se incorporaban durante la síntesis proteica, en tanto que la cadena



SINTESIS DE UN LIPIDO. Dicha síntesis depende de varios enzimas que son proteínas constitutivas de la membrana. El lípido es la fosfatidiletanolamina (PE). Los enzimas, sustratos y precursores que son solubles en el citoplasma se muestran en color; el resto de moléculas son solubles en aceite, y están confinadas en la doble capa lipídica. La síntesis comienza con la transferencia de dos ácidos grasos, a partir de moléculas donadoras (Acil CoA), al glicerol fosfato. Una molécula de citosín monofosfato (CMP) se une al glicerol, y se sustituye luego por el aminoácido serina. En la última fase de la síntesis, el PE se forma por extracción de una molécula de dióxido de carbono (CO_2) de la serina. Debido a que cuatro de los cinco enzimas necesarios para este proceso se pueden encontrar únicamente en la doble capa lipídica, dedúcese que los lípidos se forman en la misma membrana. En razón de que el enzima restante y algunos sustratos están presentes sólo en el citoplasma, la síntesis de lípidos debe ocurrir en la cara citoplasmática, de la doble capa. PE puede marcarse selectivamente con ácido trinitrobenzensulfónico (TNBS), que se une al grupo amino de la cabeza hidrofílica del lípido. (George V. Kelvin.)

en crecimiento de *G* se seguía uniendo al ribosoma. Este descubrimiento resultó ser decisivo. Si el carbohidrato se incorporaba a *G* en el lado que daba al lumen, mientras la proteína estaba creciendo aún sobre un ribosoma de la cara citoplasmática, *G* debía ser transportado necesariamente a través de la membrana durante la síntesis proteica.

Dimos con un hallazgo sugestivo al romper la membrana en un momento en que la cadena tenía más de 150 aminoácidos pero menos de 400. La proteína resultante era una forma intermedia, no observada en la naturaleza, con sólo una cadena de carbohidratos. La creación de este nuevo intermediario suministró una prueba contundente de que las dos estructuras carbohidratadas se añadían de forma secuencial a medida que el polipéptido pasaba a través de la membrana.

INCUBACION CON FOSFATO RADIACTIVO

INCUBACION CON TNBS

DISOLUCION DE LA MEMBRANA

RADIACTIVO

NO RADIACTIVO

PE MARCADO CON TNBS

PE

ORIGEN

DISOLVENTE

LUGAR DE LA SINTESIS DE LIPIDOS, que se determinó por marcado independiente del PE recién formado y de la fracción de PE que se expone en la superficie exterior de una célula bacteriana. El PE de nueva formación se marcó incubando brevemente la célula con fosfato radiactivo; éste se incorpora a todos los fosfolípidos. El PE externo se marcó con ácido trinitrobenzesulfónico (TNBS), el cual entra en combinación con todas las moléculas de PE que pueda alcanzar, si bien no puede atravesar la doble capa. Cuando se disuelve la membrana, los lípidos pueden separarse según sean o no radiactivos y estén o no marcados con ácido TNBS. (Esta última determinación se hace por cromatografía en capa fina.) Casi ninguna molécula de PE radiactivo –las sintetizadas recientemente– está marcada con ácido TNBS. Ello indica que el lípido se sintetiza solamente en la cara citoplasmática de la membrana.

La asimetría de los lípidos de la membrana difiere en sus detalles de la que caracteriza a las proteínas, pero puede aplicársele el mismo principio fundamental: la membrana puede crecer sólo por expansión. No puede romperse y abrirse para insertar nuevo material. En el caso de los fosfolípidos, esta necesidad se cubrió de forma sencilla: los lípidos se sintetizan dentro de la misma membrana.

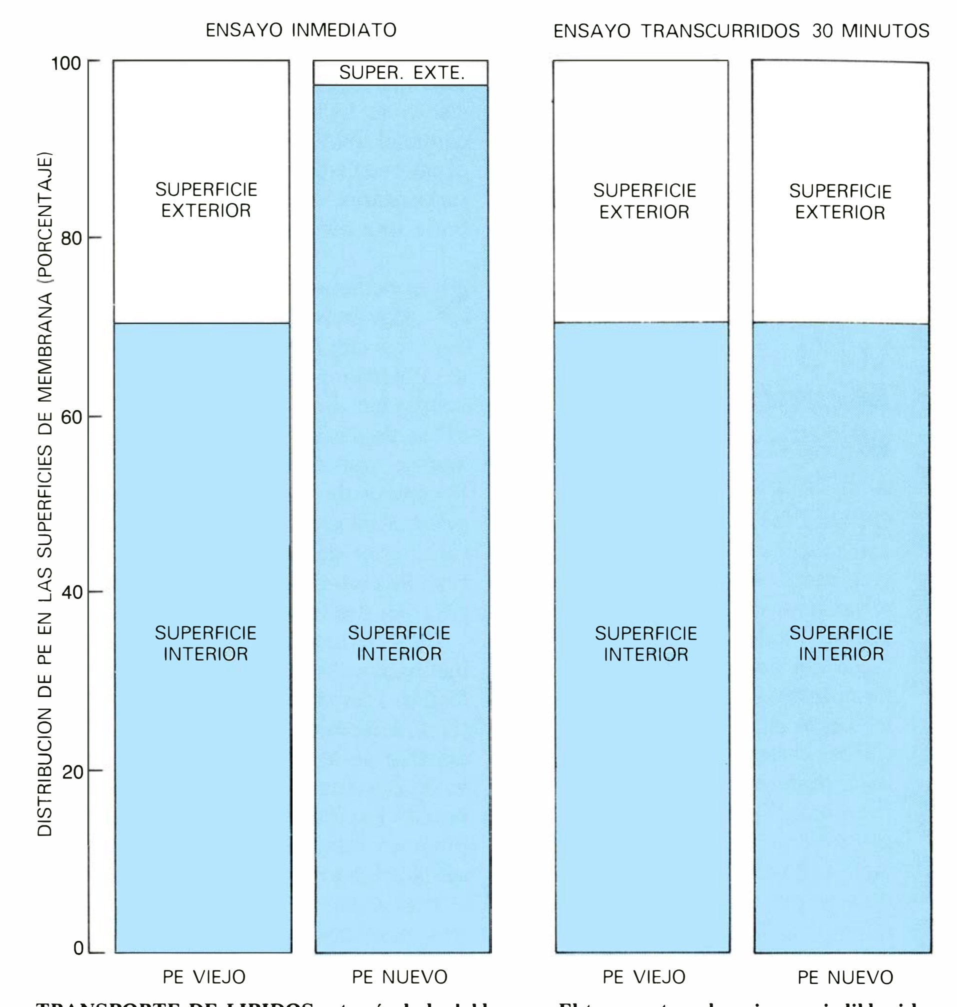
Tenemos un ejemplo en el lípido fosfatidiletanolamina, o PE, cuya síntesis en la bacteria *Escherichia coli* sigue una vía bioquímica que ha sido trazada principalmente por Eugene P. Kennedy y sus colegas, de la Harvard Medical School. La síntesis comienza con dos moléculas de ácidos grasos, que se transfieren desde un donador a una molécula de glicerol fosfato, formando el fosfolípido más primitivo, el ácido fosfatídico. Los dos ácidos grasos constituyen las colas hidrofóbicas de la molécula y la cabeza hidrofílica es el glicerol fosfato. En los pasos siguientes, un nucleótido fosforilado, CMP, se inserta en la cabeza del ácido fosfatídico; luego se reemplaza el CMP por el aminoácido serina. Por último, la serina pierde una molécula de dióxido de carbono, produciéndose PE. Cada paso de la síntesis viene catalizado por un enzima; se ha encontrado que todos los enzimas, menos uno, son proteínas constitutivas de la membrana plasmática bacteriana. Los sustratos lipídicos y los productos de las reacciones deben, por tanto, ser también componentes de la doble capa lipídica. La membrana crece por fabricación de lípidos "in situ".

La síntesis de fosfolípidos es fundamentalmente idéntica en las células eucariotas, radicando la diferencia principal en que los fenómenos no ocurren en la membrana plasmática sino en el retículo endoplasmático. De nuevo, los enzimas comprometidos son proteínas de membrana, lo cual supone que los lípidos recién formados y los intermediarios participan en la doble capa. Por consiguiente, el retículo endoplasmático constituye una verdadera fábrica de membrana: se trata del lugar donde los lípidos y la mayor parte de los componentes proteicos de las membranas se unen antes de emigrar a otras partes de la célula.

Es fácil comprender de un modo general cómo crecen las membranas lipídicas por expansión. Ahora bien, el análisis presentado más arriba soslaya el único rasgo estructural de la membrana que no hemos explicado, a saber: las dos caras de la doble capa no son iguales.

¿Se forman los nuevos lípidos sólo sobre una de las caras de la doble capa o sobre ambas? Muchos de los precursores en la síntesis lipídica y uno de los enzimas son solubles en el citoplasma; el que estas sustancias no puedan cruzar la parte central hidrofóbica de la membrana plantea el problema de que la síntesis de lípidos ocurra sólo en la superficie citoplasmática. Si ello es así, debe existir algún sistema para transportar las moléculas lipídicas hasta la superficie externa.

El mecanismo de la síntesis lipídica tiene que estar de acuerdo asimismo con la composición asimétrica de la doble capa. Esta asimetría no es absoluta, como la de las proteínas, dentro de las cuales cada especie tiene una orientación determinada, observada por cada una de las moléculas de esa especie proteica. La asimetría de la composición lipídica puede describirse de forma excelente como una predisposición estadística. No se han encontrado lípidos exclusivamente en uno u otro nivel, pero muchos parecen preferir una de las superficies. En algunas bacterias, por ejemplo, la concentración de PE en la cara citoplasmática es unas dos veces, aproximadamente, la cantidad presente en la cara externa. Otros fosfolípidos se hallan predominantemente en la capa exterior.



TRANSPORTE DE LIPIDOS a través de la doble capa. El transporte se hace imprescindible si los nuevos lípidos se forman sólo en una superficie. En la bacteria *Bacillus megaterium*, la distribución de PE en las dos capas de la membrana es asimétrica: cerca del 70 por ciento se encuentra en la cara citoplasmática. Sin embargo, cuando se mide la distribución de PE recién sintetizado, tal como se esquematiza en el experimento que se ilustra en la página anterior, se halla casi todo en la superficie interna. Para mantener la integridad de la doble capa, parte del nuevo material debe desplazarse a la superficie exterior. En la práctica, durante los 30 minutos siguientes el PE de nueva formación se equilibra con el material preexistente. Se ha sugerido que las proteínas de la membrana bacteriana en crecimiento son las que facilitan el movimiento de los lípidos a través de la doble capa, al permitirles, probablemente, alcanzar un equilibrio termodinámico. (Dibujos y esquemas de George V. Kelvin.)

Las asimetrías de la composición de lípidos han sido investigadas en el Harvard Medical School por uno de nosotros (Rothman) en colaboración con Kennedy. En varios experimentos de éstos se utilizaron membranas de *Bacillus megaterium*. Se eligió una bacteria de este tipo, llamada grampositiva, porque la presencia de una sola membrana por célula simplifica el análisis de la composición lipídica. Se optó por este organismo, en parte porque su membrana encierra grandes cantidades de PE, que constituyen alrededor del 70 por ciento del total de la membrana lipídica. La mayor parte del resto es fosfatidilglicerol, en el que el grupo etanolamina del PE se halla reemplazado por un segundo glicerol. En otras bacterias grampositivas, el fosfatidilglicerol constituye el lípido mayoritario de la membrana. La abundancia de PE en *B. megaterium* supone una ventaja experimental, ya que el PE puede marcarse fácilmente.

La distribución de PE en la membrana de *B. megaterium* se midió marcando todas las moléculas de PE expuestas en la superficie exterior, sin marcar las de la superficie citoplasmática. La proporción de PE sobre cada cara podía venir determinada por comparación entre el número de moléculas marcadas y no marcadas.

El reactivo marcador fue el ácido trinitrobenzensulfónico, o TNBS, el cual reacciona con el grupo amino de la molécula de PE, pero no se combina con el fosfatidilglicerol u otros lípidos que carecen del grupo amino. Si hay TNBS en cantidades suficientes, marcará todas las moléculas de PE a las que tenga acceso; las moléculas accesibles son, justamente, las de la superficie exterior de la membrana. El TNBS no puede marcar la cara citoplasmática de la membrana porque es soluble en agua; por tanto, tampoco puede atravesar la barrera de permeabilidad. Este sencillo e importante principio del marcado de membranas con reactivos que no penetren fue desarrollado por Mark S. Bretscher, de los laboratorios del Medical Research Council en Cambridge, quien lo aplicó a sus estudios pioneros sobre lípidos y proteínas en la membrana de los glóbulos rojos.

Finalizado el marcado, las membranas bacterianas se rompieron por disolución de los lípidos en un disolvente orgánico. Hubo que separar el PE marcado y no marcado, lo cual se hizo por la técnica de la cromatografía en capa fina. Una muestra de los lípidos extraídos se aplica al final de una fina capa de gel de sílice, y se deja ascender una mezcla de disolventes orgánicos a través del gel por capilaridad, arrastrando los lípidos extraídos. El PE marcado con TNBS es más soluble en estos disolventes que el PE sin marcar; de esta manera los lípidos marcados ascienden más por entre el gel. En cierto sentido, ambas mitades de la doble capa de fosfolípido se separaron por este método. Cuando se determinó la cantidad de cada material, se vio que sólo se había marcado alrededor del 30 por ciento del PE. Ello demostraba

que un 30 por ciento del PE, aproximadamente, estaba sobre la cara exterior de *Bacillus megaterium* y el otro 70 por ciento se hallaba dentro. Dado que la cantidad total de lípido en las dos monocapas resultaba casi igual, el otro lípido mayoritario, el fosfatidilglicerol, debía tener una distribución inversa.

Sirviéndonos de un procedimiento algo más elaborado pudimos abordar dos cuestiones: dónde se situaban los lípidos en la membrana y dónde se construían. La idea experimental consistió en emplear dos marcadores independientes: uno de ellos podía marcar sólo los lípidos de nueva formación y el otro podía distinguir los lípidos de la superficie interior de los localizados en la exterior. Se consiguió revelar así la distribución del nuevo material lipídico.

Para marcar el PE recién sintetizado incubamos las células bacterianas con fosfato inorgánico marcado con el isótopo radiactivo fósforo 32. El fosfato radiactivo se incorporó al glicerol fosfato y, de este modo, a los lípidos de membrana. La incubación se realizaba durante un minuto, aproximadamente; así sólo los lípidos que se formaban durante la exposición a la radiactividad quedaban marcados. A la suspensión de células se añadió TNBS, y se dejó que se combinara con las moléculas de PE sobre la superficie exterior de la membrana. Se disolvió la membrana a continuación, y se separaron los lípidos por cromatografía comprobando luego su radiactividad. Si ambos, el PE marcado con TNBS y el PE sin marcar, eran radiactivos, el lípido debía sintetizarse en ambas caras, ya que estas fracciones proceden de caras opuestas de la membrana. Pero descubrimos que la radiactividad sólo se detectaba en el PE no combinado con TNBS; ello indicaba que la síntesis de estos lípidos estaba restringida a la superficie citoplasmática, a la cual no llegaba el TNBS.

Este hallazgo concuerda con el origen citoplasmático de los precursores lipídicos, pero plantea también una cuestión preocupante. Si los lípidos se sintetizan en la superficie citoplasmática e inmediatamente forman parte de la capa citoplasmática, ¿cómo alcanzan siempre la otra cara de la membrana? Según se comentó más arriba, la tasa de "flip-flop" (sistema biestable) espontáneo a través de la doble capa es muy baja, de suerte que hay que esperar que un lípido determinado sufra una transición cada mes, aproximadamente. Ahora bien, una bacteria puede duplicar su masa y dividirse en menos de una hora y, en consecuencia, duplicar su superficie de membrana plasmática. Si los lípidos no pudieran desplazarse de la superficie interior a la exterior, la doble capa pronto se convertiría en una capa sencilla (monocapa) de lípido.

Otro experimento ha demostrado que los lípidos atraviesan realmente la membrana; aunque no se ha revelado cómo tiene lugar este paso. Se incubaron brevemente poblaciones bacterianas con fosfato radiactivo, pero en vez de marcarlas en seguida con TNBS, se añadió otro reactivo que inhibía la síntesis de lípidos. El inhibidor, hidroxilamina (NH_2OH), inhibe el último enzima de la secuencia que conduce al PE, a saber: el enzima que extrae una molécula de dióxido de carbono de la serina. El TNBS podría añadirse después de varios intervalos y podría mostrar de qué forma muchos lípidos sufren migraciones a través de la membrana en el período posterior al bloqueo de la síntesis de PE.

Cuando se añadió TNBS inmediatamente después del inhibidor, casi toda la radiactividad se encontró en la monocapa interna, lo cual evidenciaba de nuevo que éste era el lugar de síntesis de lípidos. Si el marcado con TNBS se retrasaba 30 minutos, la distribución de nuevo PE radiactivo resultaba similar a la del antiguo lípido no radiactivo, con cerca de un 30 por ciento de ambos en la superficie exterior. Obviamente, las nuevas moléculas de PE podían atravesar la membrana con una rapidez suficiente como para obtener su distribución asimétrica natural.

A partir de una serie de mediciones semejantes calculamos que, a temperatura fisiológica, el tiempo que se requiere para que una hornada de nuevos lípidos vuelva a alcanzar su distribución final es de alrededor de 5 minutos. Es decir, cerca de 100.000 veces menos que la tasa de "flip-flop" observada en otros sistemas de membrana; no es, pues, probable que se deba a "flip-flop" espontáneos. En consecuencia, hemos propuesto que las membranas en crecimiento incluyen proteínas que facilitan el equilibrio de los lípidos a través de la membrana, idea que lanzara por primera vez Bretscher. Estas proteínas no influirían necesariamente en la dirección del movimiento del lípido; suministrarían una mera orientación o cauce. Las proteínas podrían ser necesarias sólo cuando las membranas estén creciendo activamente, como las membranas plasmáticas en las bacterias y las membranas del retículo endoplasmático en los eucariotas; su ausencia, por otra parte, podría explicar la tasa muy baja de intercambio entre los niveles o capas de

otras membranas. Debe insistirse, de todos modos, en que no se ha comprobado la existencia de tales proteínas.

En este planteamiento de la doble capa lipídica surge una nueva complicación: si los lípidos pueden pasar libremente de una capa a la otra en una membrana en crecimiento, como la de *B. megaterium,* ¿cómo puede mantenerse la asimetría de la composición lipídica? La explicación que hoy dan numerosos investigadores, es que la asimetría no se mantiene activamente, sino que refleja un equilibrio termodinámico entre las moléculas de ambas superficies. Esta hipótesis supone que algunas especies de lípidos poseen una baja energía libre sobre una cara de la membrana y una elevada energía libre sobre la otra. La diferencia de energía no precisa ser muy elevada para explicar las asimetrías observadas, que son casi siempre modestas si se las compara con las de las proteínas. No es difícil imaginar un mecanismo que pueda desencadenar una diferencia energética de tal magnitud. Por ejemplo, si las proteínas se hallaran sólo en la superficie citoplasmática, o si el PE poseyera una mayor afinidad hacia los iones que hacia las proteínas externas de la célula, entonces cada molécula de PE podría estar más tiempo en la cara citoplasmática. De esta forma podría originarse una distribución asimétrica aunque el cambio de nivel de las moléculas lipídicas no estuviese físicamente restringido.

Esta hipótesis establece una diferencia fundamental entre las asimetrías de lípidos y proteínas. La asimetría proteica consiste en un estado de no-equilibrio resaltado durante el proceso de síntesis; si se deja que las proteínas alcancen el estado de equilibrio (mediante rotura y reconstrucción de la membrana) adquieren una distribución mucho más aleatoria. La asimetría de los lípidos, y siempre de acuerdo con esta última hipótesis, tiene su origen en un equilibrio desplazado, y no depende del modo y lugar de la síntesis lipídica; si se insertaran en la capa externa su distribución sería la misma.

Nuestras conclusiones sobre la síntesis de lípidos en las membranas bacterianas han sido recientemente ampliadas por parte de otros investigadores, quienes han señalado mecanismos similares en el retículo endoplasmático de las células eucariotas. Rosalind Coleman y Robert M. Bell, de la Duke University, han mostrado que los enzimas responsables de la síntesis de varios fosfolípidos poseen todos su actividad localizada en la cara citoplasmática de la membrana del retículo. Donald B. Zilversmit y sus colegas, de la Cornell University, han medido el movimiento a través de la membrana del retículo. Han hallado que los lípidos alcanzan un equilibrio a través de la membrana del retículo endoplasmático en cuestión de minutos, tal como ocurría en las bacterias. Por consiguiente, nuestro modelo acerca de cómo se forma la doble capa lipídica parece ser de validez general.

Los estudios descritos aquí, aun cuando se han limitado a unos pocos sistemas experimentales sencillos, han aportado luz a varios principios de la síntesis de membranas que parecen ser aplicables ampliamente. Pero otros aspectos de la formación de las membranas permanecen todavía obscuros.

Uno de estos aspectos es el transporte de material (de membrana formada) desde el retículo endoplasmático hasta otros puntos de la célula. Las membranas de orgánulos y la membrana plasmática tienen su propia composición característica de lípidos y proteínas; para mantener estas diferencias, los materiales de las membranas nuevamente formados deben clasificarse de algún modo a medida que abandonan el retículo endoplasmático o el aparato de Golgi para que lleguen a su destino correcto. Existen también proteínas específicas del retículo endoplasmático, y no está claro qué es lo que impide su desplazamiento al igual que el resto de proteínas. El supuesto mecanismo de transporte intracelular —un sistema de vesículas que emigran del retículo y se funden con otras membranas— puede servir para transportar el material de la membrana plasmática aunque se encuentre con dificultades de todo tipo. Ciertas proteínas de la membrana de las mitocondrias, por ejemplo, deben pasar las dos membranas que rodean este orgánulo.

Otra deficiencia de nuestros modelos estriba en que no describen de una manera detallada y satisfactoria el único fenómeno crucial para el emplazamiento de proteínas y lípidos: el transporte a través de la doble capa. Para las proteínas esto sólo acontece una vez, en el momento de la síntesis e inserción. Para los lípidos, hay probablemente un intercambio veloz aunque indirecto a medida que se añade a la membrana nuevo material. En ambos casos, una proteína ligada a la membrana sirve probablemente de "portero", pero sigue sin conocerse la identidad de estas proteínas y su funcionamiento. La necesidad de comprender estos fenómenos, todavía vagos e imprecisos, atraerá futuras investigaciones y será en ellos donde los estudios se manifestarán más fecundos.

La óptica de ondas guiadas

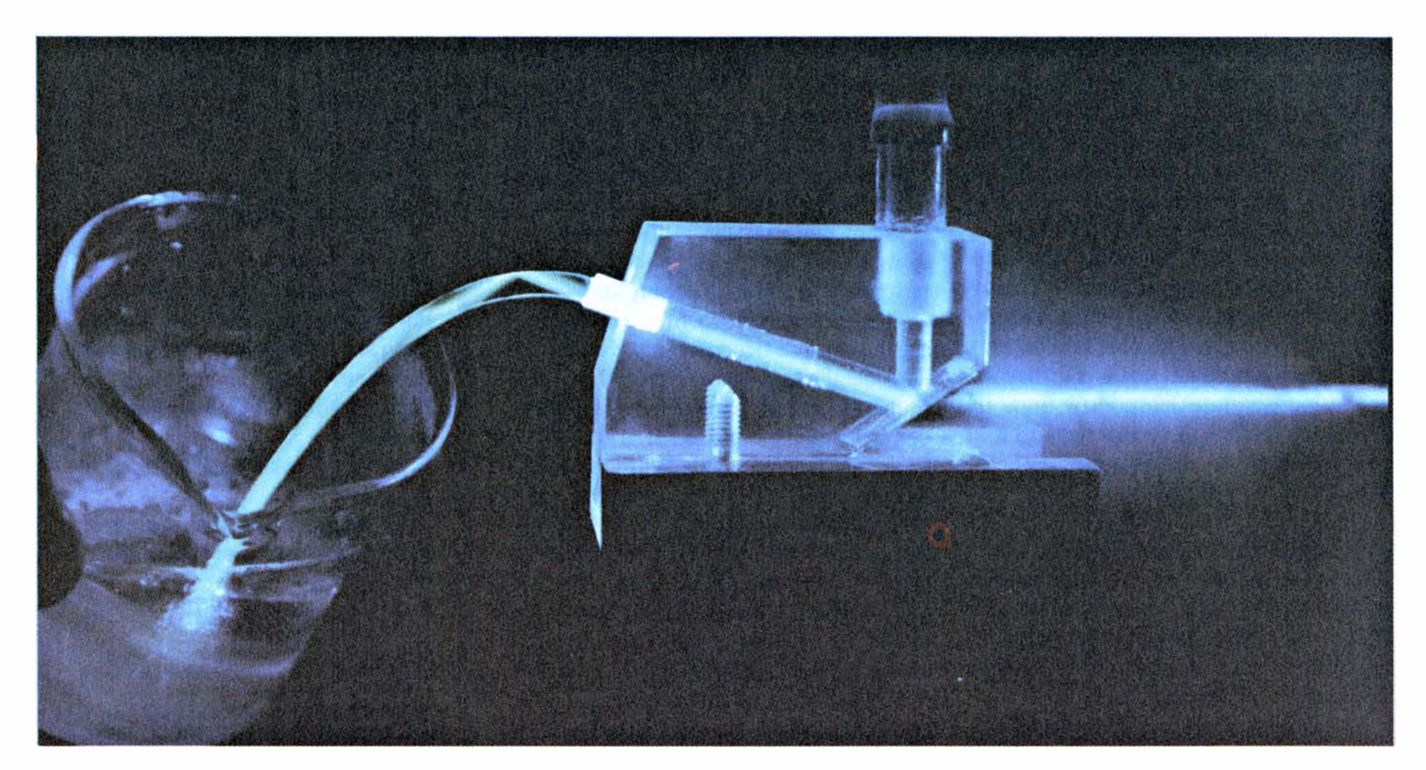
Se da cuenta del progreso alcanzado en el desarrollo de dispositivos capaces de manipular haces de un láser en finas láminas transparentes. En estas investigaciones se pretende obtener componentes pequeños y muy eficaces

Amnon Yariv

Los sistemas de comunicación convencionales por microondas están constituidos por una serie de componentes básicos. En primer lugar, estos sistemas poseen un oscilador capaz de generar ondas "portadoras" en forma altamente coherente (emisión simultánea de todas las ondas con una misma frecuencia). A continuación, dichas ondas deben ser moduladas, con el fin de que puedan transportar una determinada información variando su frecuencia o su amplitud. Finalmente, se procede a la transmisión del haz de microondas modulado, canalizándolo a lo largo de una guía de ondas metálica o propagándolo a través de la atmósfera. Dado que las ondas se atenúan al propagarse, un sistema de comunicación de este tipo para grandes distancias exigirá además la instalación de toda una serie de estaciones repetidoras intermedias. En tales estaciones, las señales de microondas serán detectadas, amplificadas, reformadas para restituirles su modulación original y transmitidas a su destino final. Al llegar a éste, el haz se detecta, se amplifica y se reforma de nuevo antes de ser procesado con el fin de recuperar la información que transporta desde el lugar de origen.

Hasta una fecha relativamente reciente hubiera sido impensable intentar disponer de un sistema de comunicación basado en las ondas luminosas que fuera comparable al antes citado. A pesar de ello, las ventajas que puede ofrecer un sistema de este tipo son evidentes: la capacidad de transportar señales de una onda electromagnética aumenta con su frecuencia y la frecuencia de las ondas luminosas resulta ser del orden de 20.000 a 200.000 veces mayor que la de las microondas.

El desarrollo del láser a principios de la década de 1960 constituyó el origen de una serie de acontecimientos que están cambiando profundamente las perspectivas de las comunicaciones ópticas. Durante estos últimos años se han producido tres avances importantes en este sentido. Uno de ellos consiste en la invención del láser de semiconductor-



PRINCIPIO DE LA GUIA DE LA LUZ, utilizado en la fabricación de guías de ondas en forma de láminas delgadas de pequeño tamaño. Dicho principio se pone de manifiesto mediante un experimento llevado a cabo por Richard Moyer y el autor de este artículo en el Instituto de Tecnología de California. La demostración original de este principio fue realizada por John Tyndall hace más de un siglo. La versión moderna consiste en hacer llegar la luz azul-verde de un láser de gas argón sobre un chorro de agua. (Tyndall trabajó con una fuente ordinaria de luz no coherente.) La diferencia entre el índice de refracción del agua y del aire hace posible que el haz resulte guiado por reflexión interna total a lo largo de la curva del chorro.

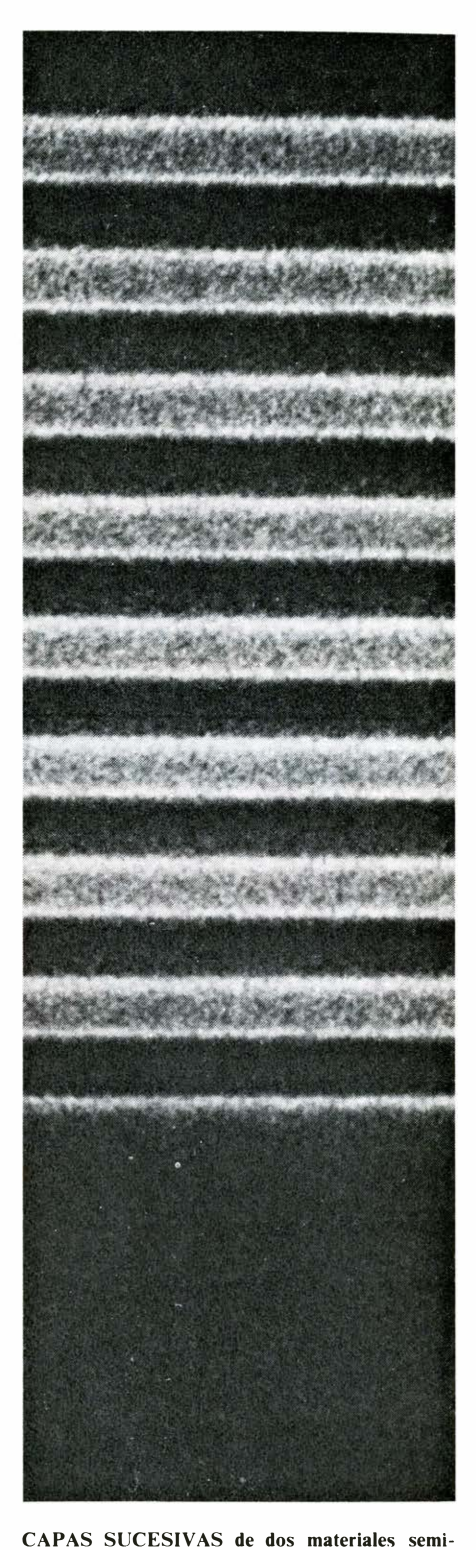
diodo, que satisface la necesidad de disponer de una fuente de ondas luminosas capaces de transportar señales codificadas; y ello con la particularidad de ser barato, eficiente y de larga vida. El desarrollo de fibras ópticas ultradelgadas y de baja pérdida, capaces de transmitir ondas luminosas a grandes distancias, constituye también una conquista importante. Por último, se está introduciendo una generación totalmente nueva de dispositivos ópticos compactos, capaces de guiar con gran precisión la luz de un láser, gracias a los cuales la obtención de microcircuitos ópticos es una realidad. Al parecer, el hecho de que los sistemas de comunicación óptica de alta capacidad y eficiencia empiecen a remplazar a los sistemas de microondas en numerosas aplicaciones es sólo cuestión de tiempo. En este artículo, vamos a analizar el estado en que se encuentra actualmente la investigación sobre un aspecto importante de los sistemas de comunicación óptica del futuro: el desarrollo de los componentes basados en el paso de la luz de un láser a través de láminas transparentes delgadas.

Cuando un haz luminoso se propaga a través de la atmósfera tiende siempre a difundirse. Este comportamiento (difracción) constituye una propiedad fundamental, no sólo de las ondas luminosas, sino de todos los tipos de ondas. En la práctica, la existencia de la difracción da lugar a que cuando un haz luminoso se dirige de un punto a otro y ambos puntos están suficientemente alejados entre sí, dicho haz resulta considerablemente atenuado (incluso en el caso de que se trate de un haz de láser altamente colimado). Si intentamos aumentar la intensidad del haz reduciendo esta "pérdida" con ayuda de una lente de focalización, produciremos una dispersión todavía mayor. En 1910, Peter W. Debye sugirió que este problema podría resolverse haciendo que la luz se propagara a través de una guía de ondas consistente en un material transparente multiestratificado.

Para comprender en qué forma un haz luminoso se confina y transmite de forma eficiente a través de una de estas guías de ondas debemos analizar con detalle lo que sucede cuando la luz se refleja en la interfase de separación entre dos medios materiales transparentes. La luz se transmite a través de cualquier medio transparente con una velocidad inferior a la de la luz en el vacío. La cuantía del retraso respecto a la velocidad de la luz en el vacío o en el aire (aproximadamente igual a 300.000 kilómetros por segundo) es una propiedad característica de cada material, conocida como índice de refracción. Por ejemplo, en un vidrio cuyo índice de refracción es 1,5, la velocidad de la luz es igual al resultado de dividir 300.000 kilómetros por segundo por 1,5, es decir, 200.000 kilómetros por segundo.

Siempre que una onda luminosa, que avanza en el seno de un material, choca contra una interfase plana que lo separa de otro material, cuyo índice de refracción es menor, la discontinuidad en el índice de refracción da lugar a que una parte de la onda se refleje en dicha superficie y retorne al primer medio, en tanto que el resto del haz resulta refractado hacia el segundo medio. Para un determinado ángulo crítico (medido en relación con la perpendicular a la interfase), la onda refractada emergerá en una dirección paralela a la interfase y, para ángulos de incidencia todavía mayores (es decir, para ángulos de incidencia rasantes), la onda refractada desaparecerá. En tal caso, la luz no se transmite a través del segundo material, y se dice que la onda incidente sufre una reflexión interna total. El fenómeno de la reflexión interna total solamente tiene lugar en el caso de que el índice de refracción del material en el que se propaga la luz sea mayor que el del material transparente adyacente. Por ejemplo, cuando un haz luminoso se propaga a través de un bloque de vidrio cuyo índice de refracción vale 1,5, el ángulo crítico de incidencia para la reflexión interna total por una interfase vidrio-aire es igual a 41,8 grados.

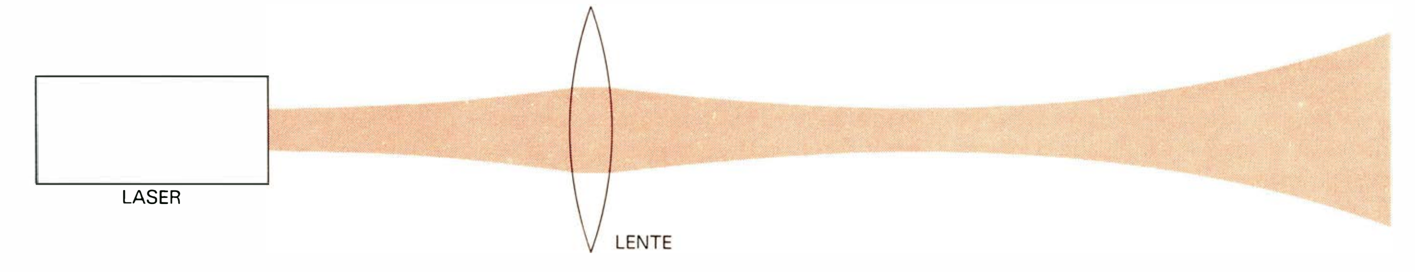
En general, el tipo más sencillo de guía de ondas transparente que existe se obtiene intercalando un material de índice de refracción elevado entre dos láminas de otros materiales cuyos índices de refracción sean menores (en forma de sandwich). En una de estas guías de ondas un rayo de luz avanza en zigzag en el seno del primero de los materiales citados, sufriendo una serie de reflexiones totales sucesivas en las interfases superior e inferior. La única condición a satisfacer consiste en que el ángulo de incidencia del rayo luminoso para ambas interfases tiene que ser mayor que el ángulo crítico para una reflexión interna total. Rodeando, por sus lados, un material transparente a la luz por un segundo material cuyo índice de refracción sea menor que el del anterior se puede obtener un canal eficiente para la transmisión de ondas. La sección transversal de dicho canal puede ser rectangular, como sucede en el caso de una guía de ondas en forma de lámina delgada, o circular, como es el caso de una fibra óptica.



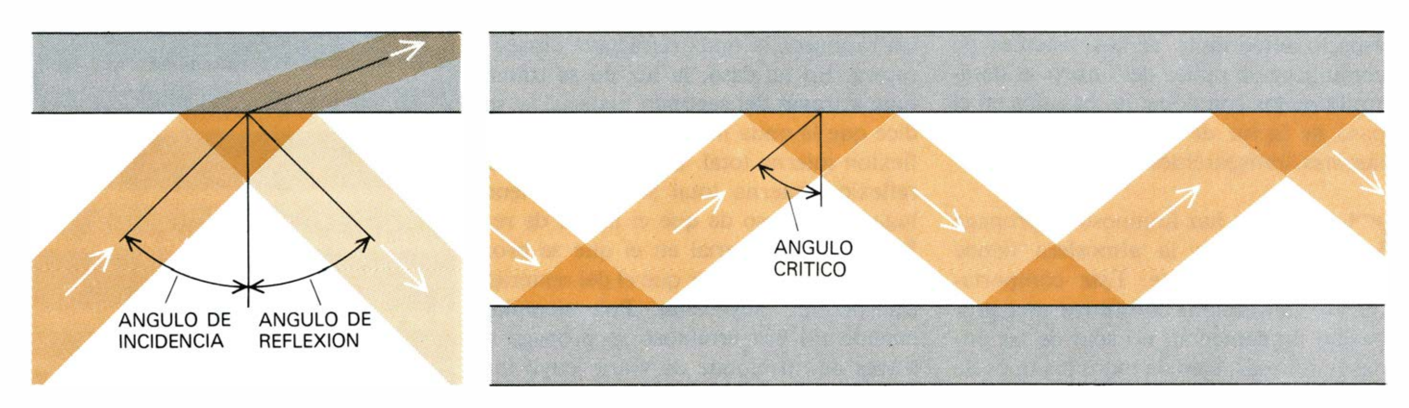
CAPAS SUCESIVAS de dos materiales semiconductores transparentes, arseniuro de galio y arseniuro de galio y aluminio. Esta fotografía se obtuvo con la ayuda de un microscopio electrónico. Dichas capas tienen un espesor de tan sólo algunos micrometros y han sido fabricadas mediante una nueva técnica de epitaxia de haces moleculares de alta precisión. La estructura de cristal individual que resulta se puede utilizar para atrapar las ondas luminosas que penetran por la capa más superior desde el otro lado, confinándolas en el seno de las primeras capas del conjunto a medida que dichas ondas avanzan en una dirección paralela a la interfase que separa la capa más superficial y el aire ambiente. Esta forma peculiar de propagación luminosa se observó, por vez primera, por el autor y Pochi A. Yeh, en el Instituto de Tecnología de California (Cal Tech), y por Alfred Y. Cho, en los Laboratorios Bell.

Aun cuando la luz se refleja totalmente en las interfases de estas guías de ondas, una pequeña fracción de la energía electromagnética que se propaga se transmite fuera del canal de guía. En general, la intensidad del campo electromagnético es máxima en el centro de la guía y disminuye gradualmente a medida que nos alejamos de dicho centro. El hecho de que la energía transmitida no esté totalmente confinada en el canal en cuestión juega un importante papel en muchas aplicaciones.

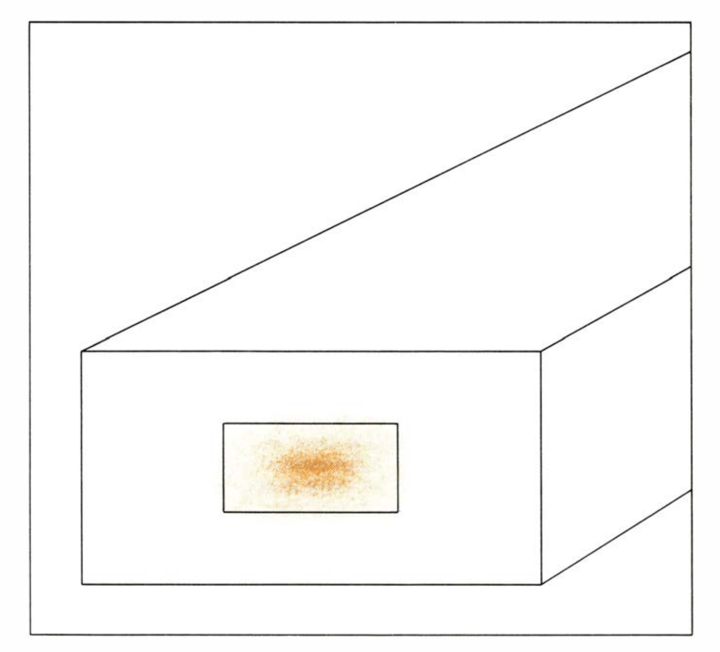
¿Cómo se construyen las guías de ondas de láminas delgadas transparentes? Una clase importante de estas guías se fabrica produciendo una serie sucesiva de capas cristalinas de materiales semiconductores a partir de su fase líquida. En esta técnica se parte generalmente de un sustrato muy pulido de al-

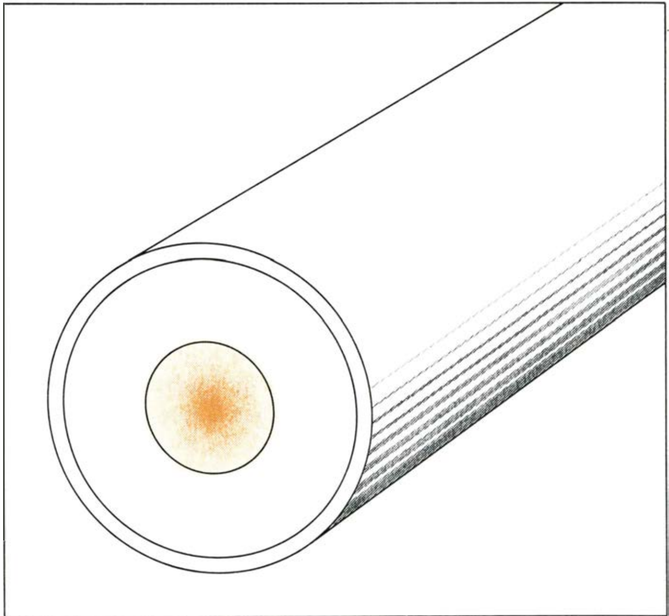


DIFRACCION de un haz de láser al propagarse en la atmósfera. Como consecuencia de este efecto (cuya cuantía se ha exagerado en esta figura), la intensidad del haz resultará atenuada proporcionalmente al cuadrado de la distancia. Por ejemplo, cuando un haz de láser cuyo diámetro inicial sea igual a un centímetro recorre la distancia de un kilómetro, su diámetro resultará ser del orden de veinte centímetros. De acuerdo con ello, la intensidad del haz se habrá reducido en un factor de 400. El intento de focalizar dicho haz con ayuda de una lente produciría una dispersión todavía mayor.



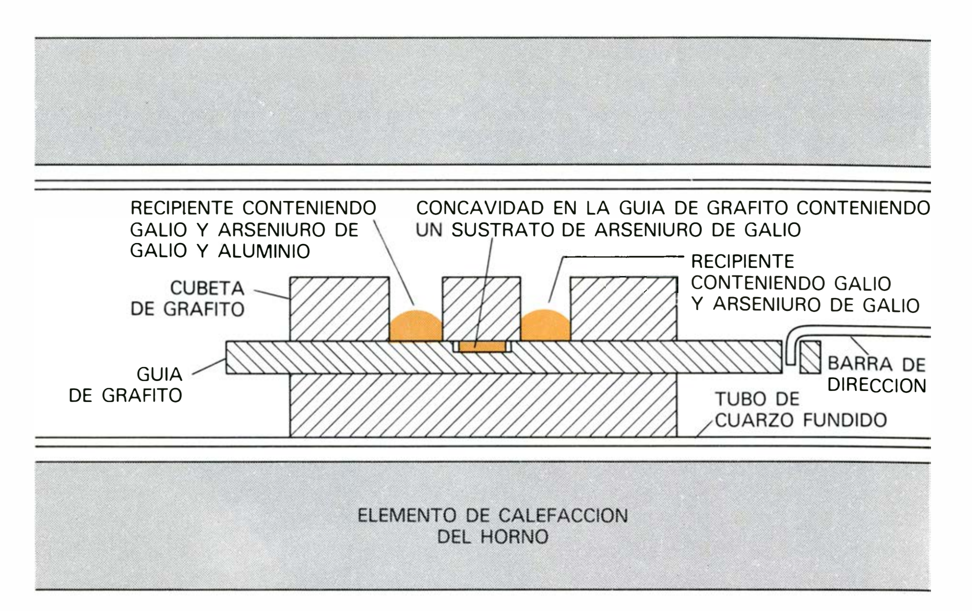
REFLEXION de un haz de láser en la interfase plana que forma el material transparente en el que se propaga dicho haz con otro medio también transparente cuyo índice de refracción es menor que el del primero *(izquierda)*. Una parte del haz es devuelta por reflexión, en tanto que el resto penetra en el segundo medio por refracción. Si el ángulo de incidencia supera cierto valor crítico, el haz sufre una reflexión total interna *(derecha)*.





CANALIZACION de un haz de láser; puede obtenerse rodeando completamente el material conductor de la luz por un segundo material cuyo índice de refracción sea menor que el del primero. La sección transversal de este canal o guía de ondas puede ser rectangular *(izquierda)* o circular *(derecha)*. En cualquiera de ambos casos, una pequeña fracción de la energía electromagnética que se propaga *(color)* se transmitirá al exterior del canal de guía. La estructura de la derecha es típica de las denominadas fibras ópticas, que consisten en un núcleo interno de vidrio revestido por otro vidrio.

gún semiconductor cristalino transparente, tal como el arseniuro de galio (GaAs). A continuación, se procede a depositar sobre él una capa cristalina delgada de arseniuro de galio y aluminio ($Ga_{1-x}Al_xAs$), un semiconductor relacionado íntimamente con el anterior, en el que una cierta fracción, x, de los átomos de galio han sido remplazados por átomos de aluminio. Esta capa se cubre por una capa delgada de arseniuro de galio puro que, a su vez, está cubierta por otra capa de arseniuro de galio y aluminio. El índice de refracción del arseniuro de galio y aluminio es menor que el del arseniuro de galio y, en consecuencia, la luz se puede guiar en el seno de la capa de arseniuro de galio.



FABRICACION de una guía de ondas de láminas semiconductoras delgadas, susceptible de ser utilizada en la investigación en el campo de las comunicaciones mediante ondas luminosas. La fabricación se lleva a cabo en el dispositivo especialmente diseñado para dicho fin que se representa en esta ilustración. Las diferentes capas de la correspondiente estructura se hacen crecer epitaxialmente (es decir, en capas cristalinas sucesivas idénticas), poniendo en contacto un sustrato cristalino de arseniuro de galio con unos recipientes de galio fundido saturados con arseniuro de galio (GaAs) o arseniuro de galio y aluminio ($Ga_{1-x}Al_xAs$), un compuesto en el que una cierta fracción (x) de los átomos de galio ha sido remplazada por átomos de aluminio, dando lugar a que el segundo de los materiales citados tenga un índice de refracción menor que el del primero. El sustrato se mueve por medio de unos rodillos situados bajo los sucesivos recipientes. El crecimiento del cristal se inicia disminuyendo adecuadamente la temperatura. El espesor de cada capa se controla ajustando la velocidad de enfriamiento del horno y el tiempo durante el que se mantiene el contacto con los citados recipientes. Algunas de las guías de ondas que han sido desarrolladas recientemente están constituidas por hasta ocho series de capas y, en consecuencia, su fabricación exige una instalación con ocho recipientes diferentes. (Gabor Kiss.)

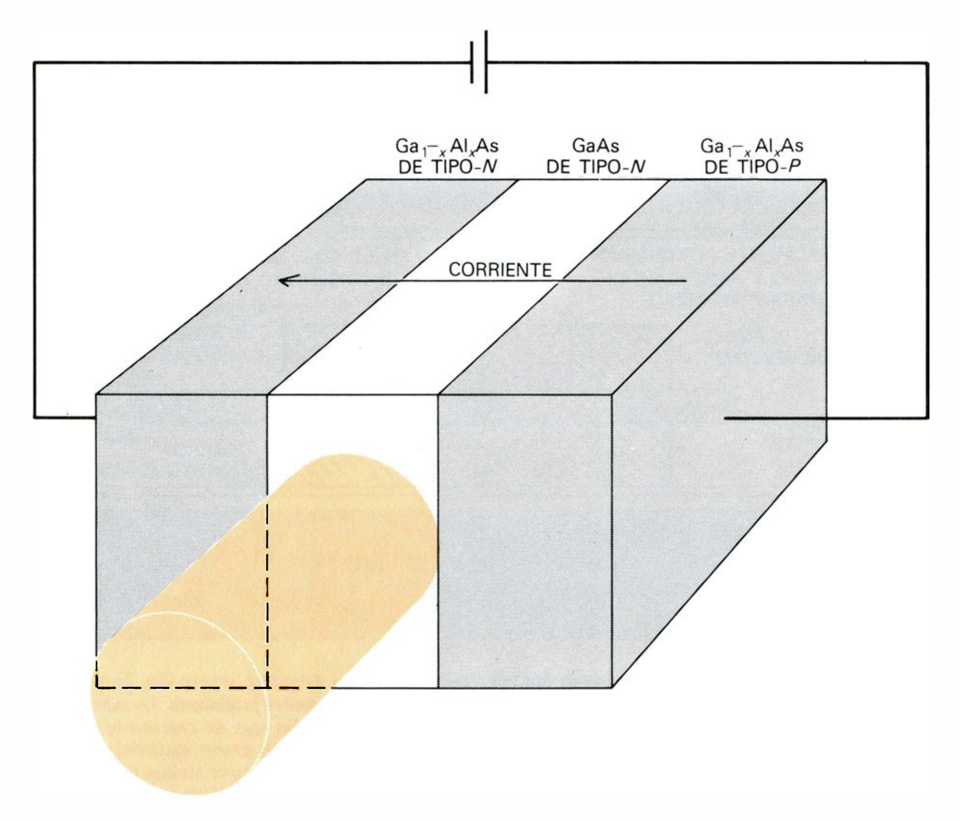
Las diferentes capas que constituyen la estructura en cuestión se pueden hacer crecer poniendo en contacto el sustrato de arseniuro de galio con unos recipientes de galio fundido saturados con uno u otro compuesto y disminuyendo adecuadamente la temperatura para iniciar el crecimiento del cristal. Dado que el arseniuro de galio y el arseniuro de galio y aluminio poseen la misma estructura cristalina, la estructura multiestratiforme resultante consiste de hecho en un cristal único sin solución de continuidad.

Esta estructura presenta otra propiedad importante, según la cual el comportamiento eléctrico de las capas semiconductoras se puede controlar individualmente añadiendo átomos de impurezas a la vasija de fusión. La adición de átomos dadores de electrones, tales como el estaño o el teluro, da como resultado un material de tipo *n* en el cual la corriente se conduce gracias a los portadores negativos de carga (los electrones extra donados). La adición de átomos aceptores de electrones, tales como el zinc, da como resultado un material de tipo *p*, en el cual la corriente se conduce gracias a los portadores de cargas positivas (las vacantes de electrones móviles o "huecos"). La aniquilación mutua de electrones y huecos cerca de la interfase de una unión de semiconductores *p–n* conduce a la emisión de ondas luminosas. En el año 1963, Rogerio C. Leite y yo, trabajando en los Laboratorios Bell, encontramos que la luz emitida por este dispositivo resultaba canalizada por la guía de ondas dieléctrica en las proximidades de la unión *p–n*. Esta observación ayudó a despertar el interés hacia la investigación en el campo de la óptica de ondas guiadas.

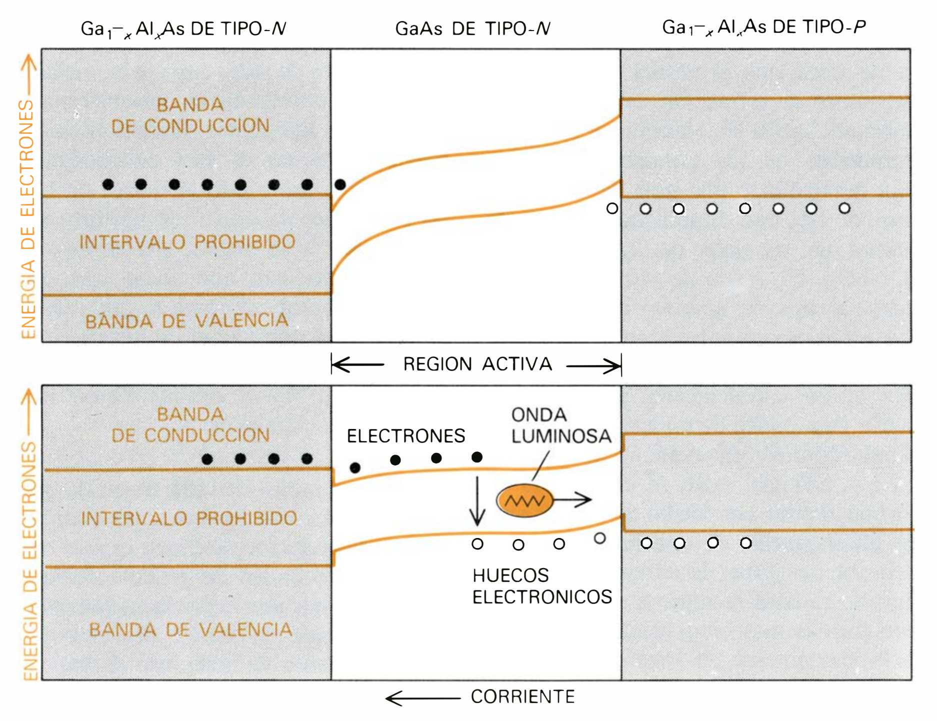
Las guías de ondas transparentes se pueden obtener también cambiando la composición química de un material mediante la introducción de un componente traza que produzca la alteración del índice de refracción del material en cuestión. Jacob M. Hammer y sus colaboradores, de los Laboratorios RCA, han desarrollado una guía de ondas común de este tipo difundiendo átomos de titanio en niobiato de litio cristalino ($LiNbO_3$). El índice de refracción de la delgada capa de niobiato de litio en la que están presentes los átomos de titanio es mayor que la del material que le rodea y, en consecuencia, puede servir como guía óptica de ondas. El titanio se puede difundir en el seno del material base y, además, existe la posibilidad de formar figuras por medio de una técnica de fotolitografía; de este modo, se pueden obtener guías de ondas no sólo del tipo de canales rectilíneos sino también con formas más complicadas.

Probablemente, el láser semiconductor-diodo es el único tipo de láser que puede tener un fuerte impacto en el mercado en un futuro próximo. Esta circunstancia se debe fundamentalmente a su simplicidad y pequeño tamaño. En un láser de este tipo, la energía de una corriente eléctrica de baja intensidad que atraviesa el cristal semiconductor multiestratificado se convierte directamente en luz láser. El tamaño extremadamente pequeño de este dispositivo (menos de un milímetro de lado) supone la posibilidad de su utilización en microcircuitos electrónicos integrados. Dichos circuitos pueden gobernar el láser directamente. Este factor, junto con el hecho de que las longitudes de onda que produce corresponden a la región del infrarrojo próximo, explican que dicho láser resulte muy apropiado para la generación de las ondas portadoras de los sistemas de comunicación por ondas luminosas que utilizan fibras ópticas como elemento de transmisión.

Desde el punto de vista de un diseñador de guías de ondas, un láser del tipo semiconductor-diodo no es más que una guía de ondas de película delgada compuesta por una capa de arseniuro de galio colocada en forma de sandwich entre dos capas de arseniuro de galio y aluminio. Para comprender la forma en que la luz que se propaga en el seno de la capa más interna de arseniuro de galio se puede amplificar por medio de una corriente eléctrica que atraviese dichas capas, es necesario exponer primero de qué forma el "doping" (dopado, adición de impurezas) controlado de los estratos con átomos de impurezas afecta a la dis-



EL LASER DE SEMICONDUCTOR-DIODO consiste esencialmente en una lámina delgada de arseniuro de galio situada entre dos láminas de arseniuro de galio y aluminio, formando una especie de sandwich. La aplicación de un voltaje externo a través de este apilamiento da lugar a la aparición de una corriente eléctrica que se mueve según una dirección perpendicular a las láminas *(de derecha a izquierda)*. La corriente se transporta gracias a los electrones adicionales que se mueven hacia el interior a partir de la capa cargada negativamente (dopada *n*, esto es, con impurezas de tipo *n*) y a las vacantes electrónicas o huecos que se mueven también hacia el interior a partir de la capa cargada positivamente (dopada *p*, es decir, con impurezas de tipo *p*). Los electrones y los huecos se recombinan en la capa central de arseniuro de galio (GaAs), emitiendo su energía en exceso en forma de luz visible.



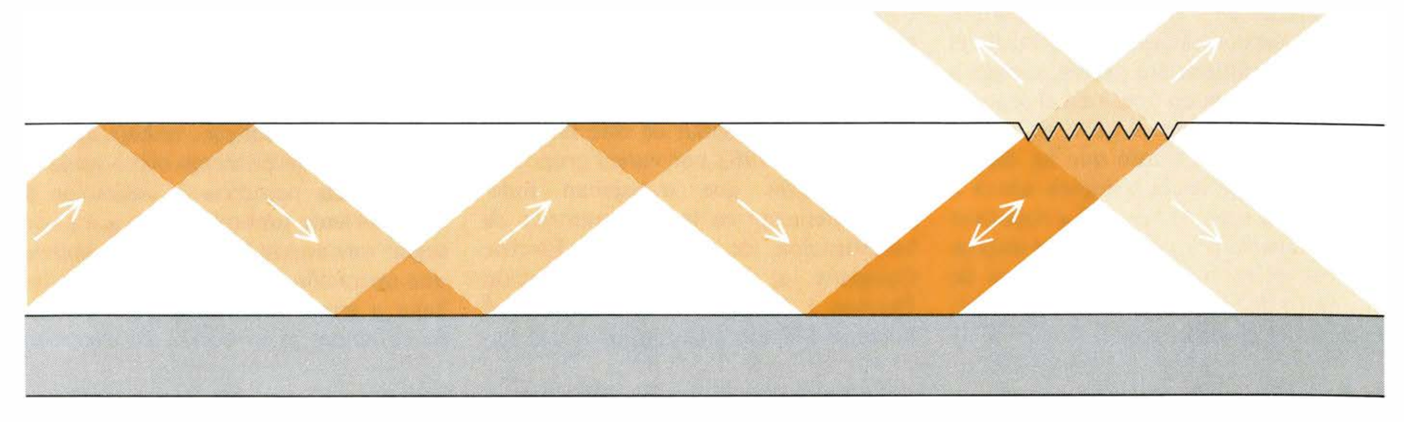
ESQUEMAS DE NIVELES ENERGETICOS, que muestran la relación existente entre la energía de los electrones y los huecos, por una parte, y su localización en la estructura estratificada de un láser semiconductor-diodo, por otra. En ausencia de un voltaje aplicado *(figura superior)*, los electrones *(puntos negros)* y los agujeros *(círculos blancos)* tienen la misma energía y se encuentran separados espacialmente. Si se aplica un voltaje "de avance" adecuado *(figura inferior)*, los electrones y los huecos se separan energéticamente, pero, en cambio, son capaces de cruzar la capa central de arseniuro de galio. En tal situación, los electrones pueden saltar a través del "intervalo prohibido" de energías, situado entre la banda de conducción y la banda de valencia, recombinándose con los huecos que están en una banda de energía más baja. Cada una de estas transiciones da lugar a la emisión de un fotón de radiación con una longitud de onda de 0,85 micrometros. (Esquemas ilustrativos de Gabor Kiss.)

tribución espacial de los electrones y a sus energías.

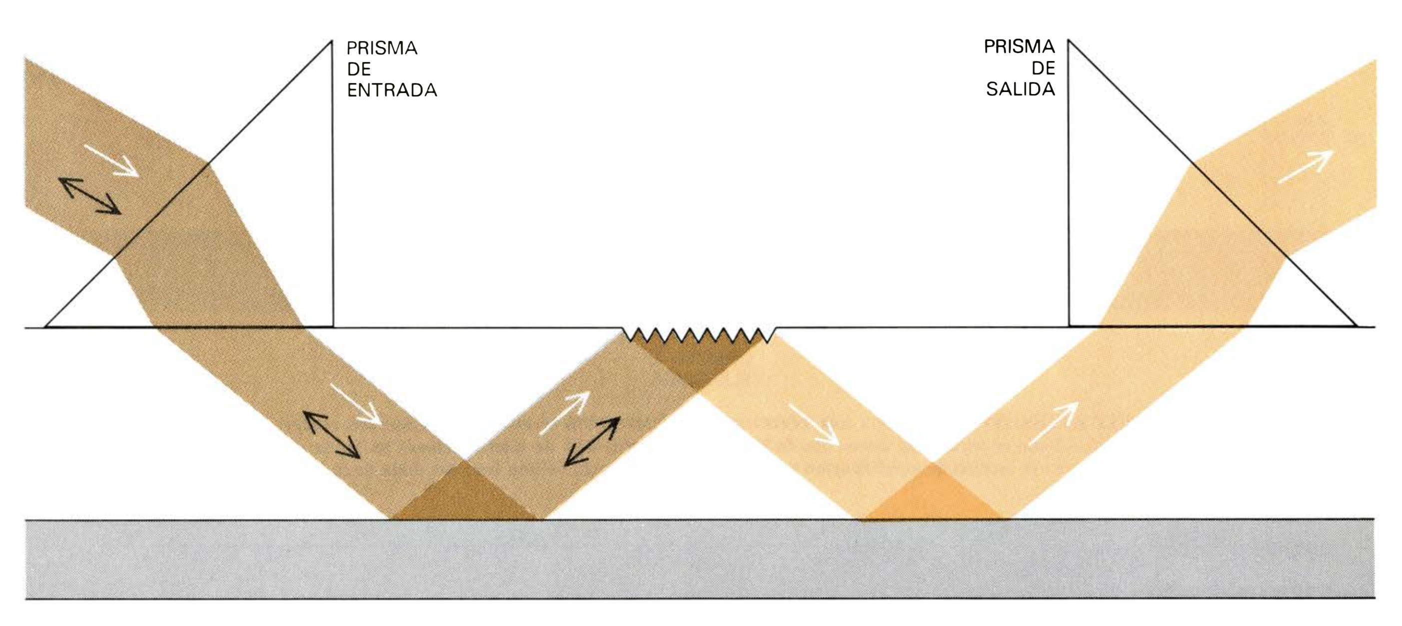
En ausencia de un voltaje aplicado, los electrones de conducción en el arseniuro de galio y aluminio de tipo *n* sobre una cara de la capa de arseniuro de galio y las vacantes electrónicas en el arseniuro de galio y aluminio de tipo *p* sobre la otra cara tienen aproximadamente la misma energía. Las altas barreras de potencial que existen en las interfases impiden que tanto los electrones como los huecos penetren en la capa interior de arseniuro de galio. La aplicación de una elevada tensión "de avance" a esta estructura produce un aumento de la energía de los electrones situados junto a la cara de tipo *n* en relación con la energía de los huecos situados junto a la cara de tipo *p*. Dicha tensión origina una disminución de las barreras de potencial anteriormente citadas y, en consecuencia, permite que los electrones de la región de tipo *n* y los huecos de la región de tipo *p* se desplacen hacia la capa central de arseniuro de galio.

Como resultado de este proceso, los electrones con una cierta energía comparten un mismo espacio con los huecos cuya energía es inferior. En esa situación, los electrones pueden ser estimulados por la presencia de luz de longitud de onda adecuada, provocando su transición a través del intervalo existente entre la banda de conducción de energía más elevada y los estados vacíos (los huecos) situados en la banda de valencia, de energía más baja. En cada una de estas transiciones se emite un fotón o cuanto de luz. Puesto que la energía de la luz estimulada se suma coherentemente a la de la luz que provoca el estímulo, el proceso transcurre en la misma forma que el mecanismo de amplificación del láser. El otro criterio principal para la acción de un láser –la provisión de un mecanismo de realimentación– se satisface gracias a la reflexión de las ondas luminosas emitidas en las superficies límites del cristal. La longitud de onda de la radiación emitida viene determinada por el valor del intervalo energético que separa a los electrones de los agujeros o huecos; por ejemplo, en el arseniuro de galio, la longitud de onda resultante es de 0,85 micrometros.

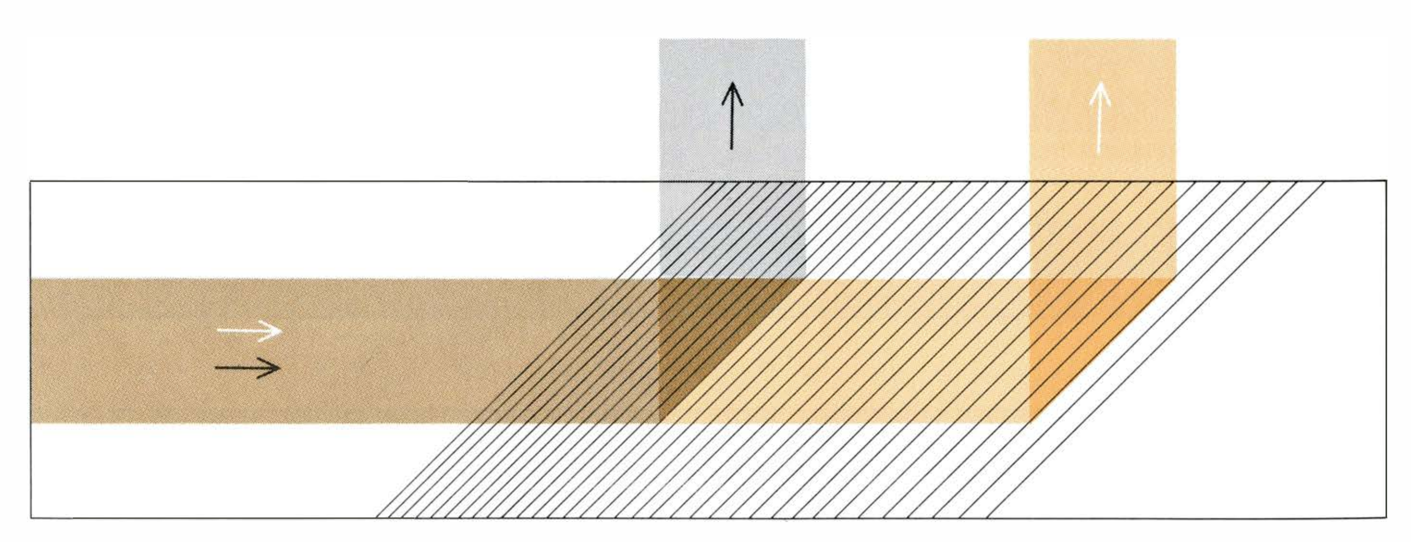
Todo este proceso se puede resumir siguiendo el recorrido de un electrón individual. En primer lugar, el electrón se extrae de una capa de tipo *p* y entra en el conductor externo dejando un hueco en la banda de valencia del material. A continuación, la energía del electrón en cuestión aumenta al atravesar la fuente de tensión (por ejemplo,



SECCION ONDULADA de una de las interfases de una guía de ondas de láminas delgadas, que refleja y refracta la luz de un láser en una serie de direcciones discretas. Estas direcciones dependen de la longitud de onda de la luz y del período de las ondulaciones. Se muestran cuatro de ellas.



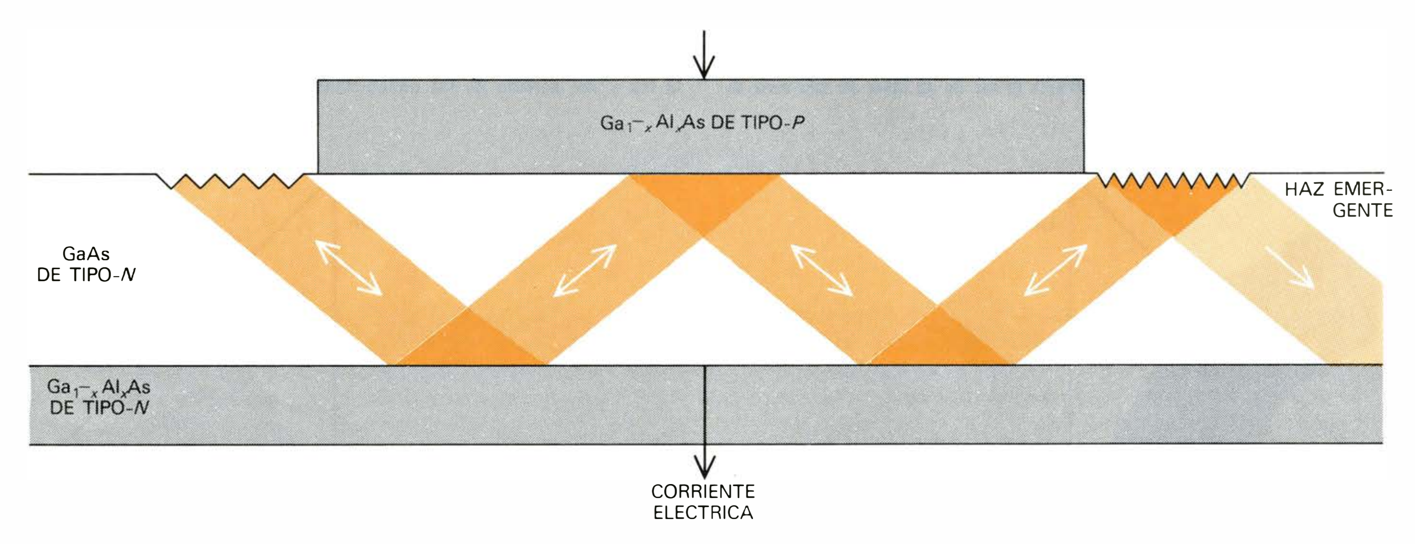
DOS HACES DE LASER coincidentes y con longitudes de onda diferentes se pueden hacer entrar con ayuda de un prisma en una guía de ondas de láminas delgadas onduladas *(izquierda).* Si el período de las ondulaciones es igual a la semilongitud de onda de uno de los haces o a un múltiplo de este valor (condición de Bragg), la ondulación actuará como un espejo, reflejando hacia atrás el *(gris).* El otro haz *(color)* no resultará afectado por las ondulaciones y emergerá prácticamente inalterado por el prisma de salida. En consecuencia, este dispositivo se comporta como un filtro de bandas.



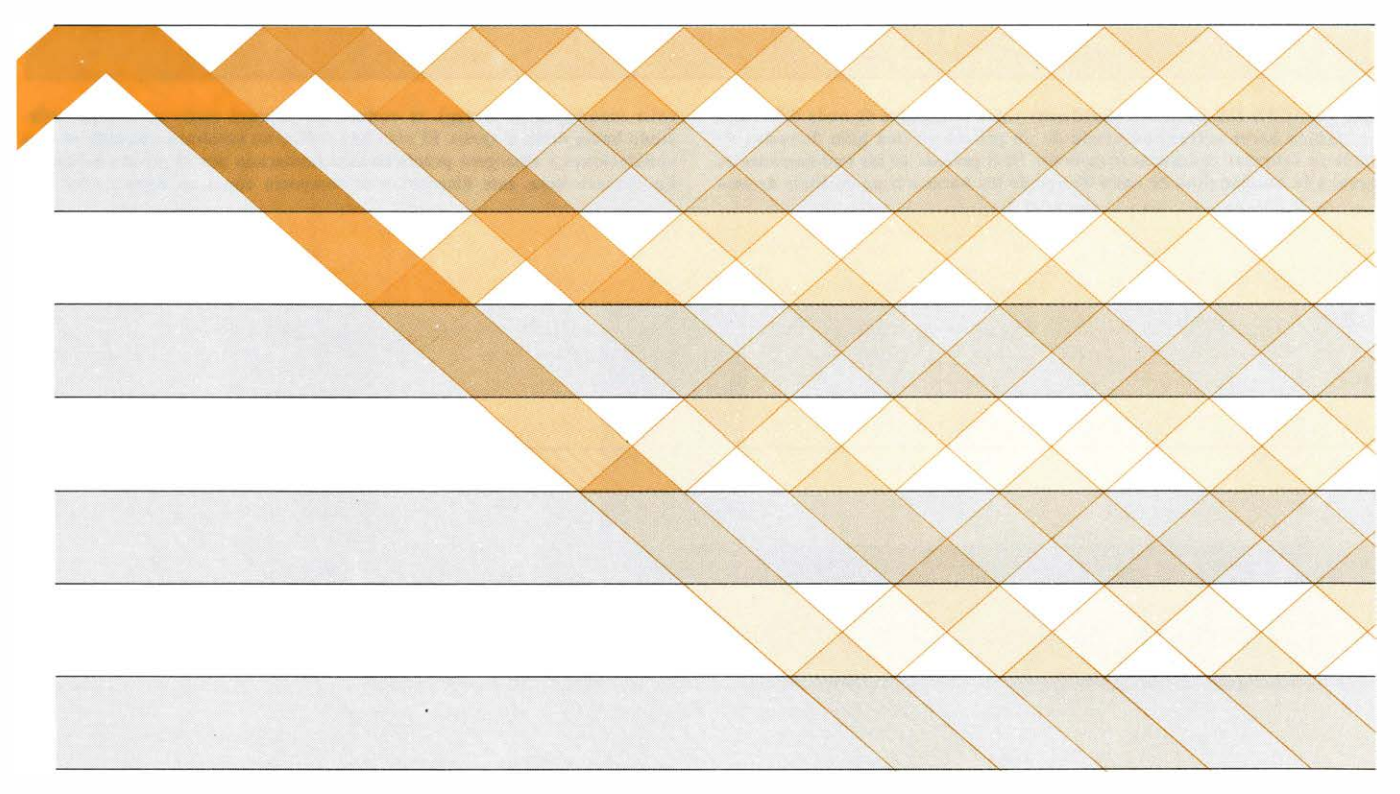
OTRA SOLUCION al problema de separar dos haces de láser coincidentes y con distintas longitudes de onda en una guía de ondas ondulada se puso a punto recientemente por los miembros del grupo de investigación del autor en el Cal Tech. En este caso, el período de la ondulación es variable y los haces coincidentes entran en la guía formando un ángulo de 45 grados con dichas ondulaciones (tal como aparece aquí). La luz cuyas longitudes de onda satisfacen la condición de Bragg será reflejada según una dirección perpendicular respecto a los haces incidentes. Puesto que el período de las ondulaciones varía a lo largo de la dirección de propagación, las diferentes longitudes de onda *(gris y color)* se reflejarán en lugares diferentes.

una batería) en una cuantía proporcional a la tensión aplicada. Más tarde, el electrón se difunde sin perder energía a través de la capa de tipo *n* en el seno de la capa de arseniuro de galio activo. Al mismo tiempo, el hueco que ha dejado detrás se dirige hacia la región activa a través de la capa de tipo *p*. La radiación láser, atrapada por una reflexión interna total en el seno de la misma capa de arseniuro de galio, estimula la recombinación del electrón con el hueco de la banda de valencia; la diferencia entre las correspondientes energías se imparte a la onda que se propaga y, como resultado, dicha onda resulta amplificada.

La versión original de este láser fue inventada en 1962 por varios grupos de investigadores que trabajaban independientemente en los Laboratorios de Investigación de la General Electric Company, el Centro de Investigación Thomas J. Watson de la International Business Machines Corporation y el Laboratorio Lincoln del Instituto de Tecnología de Massachusetts. Los primeros láseres de semiconductor-diodo estaban constituidos totalmente por arseniuro de galio. En un principio, la baja efectividad de estos dispositivos como guías de ondas hacía necesaria la utilización de una corriente umbral intensa para conseguir una acción láser y, para mantener una operación continua, era preciso trabajar a temperatura muy baja. La idea de combinar la inyección de electrones



INCORPORANDO DOS SECCIONES ONDULADAS a una estructura de láminas delgadas semiconductor-diodo se obtiene un nuevo tipo de láser. Se montan en cada uno de los lados de la sección de amplificación central y desempeñan el mismo papel que los espejos reflectores de los láseres. La reflectividad de tales secciones se puede ajustar de suerte que permita la emergencia de una fracción dada de la luz en el haz de salida *(color claro)*.

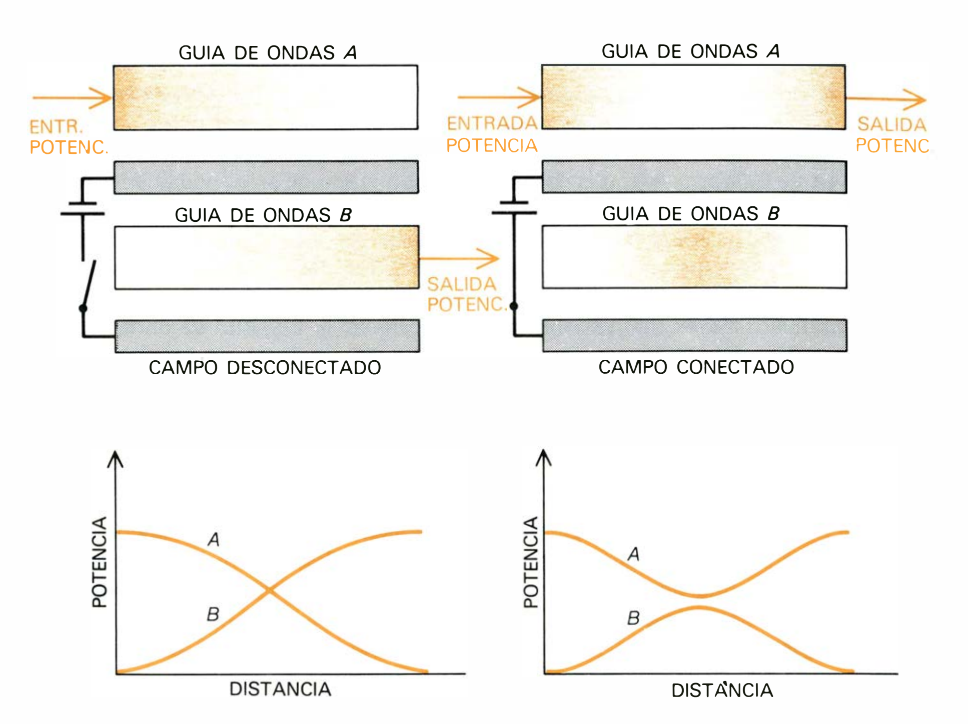


PROPAGACION DE ONDAS LUMINOSAS en las proximidades de la interfase situada entre un material semiconductor estratificado y un medio homogéneo (en este caso, aire). La estructura estratiforme es la misma que la reproducida en la fotografía de la página en que da comienzo este artículo y que ha sido obtenida mediante un microscopio electrónico. El haz de láser penetra por la izquierda en la capa más superior de arseniuro de galio (GaAs) y queda confinado en las primeras capas de este dispositivo en virtud de la reflexión interna total que se produce en la interfase que forma dicho material con el aire, así como por las sucesivas reflexiones de Bragg en fase que tienen lugar en las interfases que forman las diferentes capas del arseniuro de galio y del arseniuro de galio y aluminio. En el presente diagrama la intensidad se anula prácticamente al cabo de siete capas.

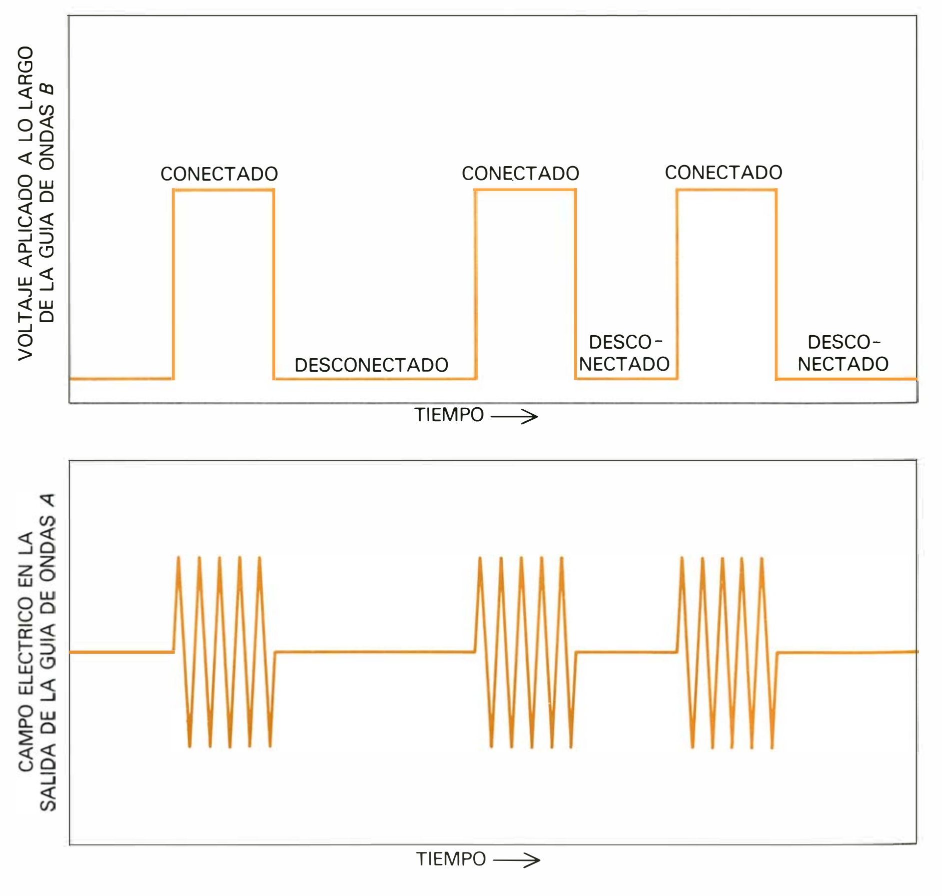
con una estructura de guía de ondas estratificada, colocando una lámina de arseniuro de galio en forma de sandwich, entre dos láminas de arseniuro de galio y aluminio, fue lanzada en 1969 por dos investigadores rusos, Zhores J. Alferov y Rudolph Kazarinov. Esta sugerencia constituye la base de los láseres de semiconductor-diodo que se caracterizan por el bajo valor de su voltaje umbral y que han alcanzado una gran difusión en la actualidad.

Numerosas aplicaciones en el campo de la óptica de ondas guiadas exigen la incorporación de una perturbación periódica en la estructura de la guía de ondas. La forma de perturbación más corriente consiste en la utilización de una sección ondulada en una o más de las interfases de la guía óptica de ondas. Vamos a considerar lo que sucede cuando una onda luminosa que se propaga en una guía de ondas se encuentra con una irregularidad de este tipo. Cada uno de los escalones de la sección ondulada difunde la luz reflejada y refractada en todas direcciones. Sin embargo, pueden existir algunas direcciones concretas para las cuales se sume la radiación difundida por todos los escalones. En cada caso particular, estas direcciones dependerán de la longitud de onda de la luz utilizada y del período de las ondulaciones (es decir, de la distancia entre dos escalones adyacentes).

Si el período de las ondulaciones es igual a la semilongitud de onda de la luz que se propaga por la guía de ondas o a un múltiplo de dicho valor (situación que se conoce con el nombre de condición de Bragg), la luz se reflejará exactamente hacia atrás, propagándose en sentido opuesto al de la onda incidente. En otras palabras, la sección ondulada se comporta como un espejo. Una sección ondulada de pequeña longitud en una guía de ondas reflejará fuertemente aquellas ondas luminosas que satisfagan la condición de Bragg y dejará pasar con pérdidas muy pequeñas a las otras ondas. Esta propiedad se puede utilizar como base de un nuevo tipo de filtros ópticos (dispositivos capaces de seleccionar o rechazar ciertas bandas de longitudes de ondas). En uno de estos filtros la onda incidente se propaga a lo largo de la guía de ondas según una dirección perpendicular a la ondulación. Puesto que en este caso el filtro rechaza la estrecha banda de longitudes de onda que satisface la condición de Bragg, se le conoce como filtro de rechazo de banda. El primer filtro de este tipo se utilizó con éxito en los Laboratorios Bell por parte de Dale C. Flanders, H. W. Kogelnik y Charles V. Shank en el año 1974.



DISPOSITIVO DE ACOPLAMIENTO DIRECCIONAL DE ONDAS consistente en dos guías de ondas transparentes y adyacentes en los que las porciones más externas de las ondas luminosas que se propagan por su interior se solapan entre sí. Con el campo de control desconectado, una onda luminosa introducida en la guía *A* se acoplará coherente y totalmente en la guía de ondas *B (diagrama de la parte superior izquierda)*. La aplicación de un campo eléctrico de control a través de la guía de ondas *B* modificará la velocidad de las ondas en esta guía y, en consecuencia, alterará el intercambio de potencia *(diagrama de la parte superior derecha)*. En el centro de la sección de interacción efectiva de este dispositivo, la mitad de la potencia de la guía de ondas *A* se habrá transferido a la guía de ondas *B*, pero en el extremo de salida toda la potencia habrá retornado a la guía de ondas *A*. Los gráficos subyacentes representan la intensidad de la onda luminosa en cada guía de ondas en función de la distancia.



EL ACOPLAMIENTO DE ONDAS DIRECCIONAL se puede utilizar para imprimir información por impulsos binarios en las ondas luminosas emergentes. Aplicando un tren de impulsos de voltaje codificados a lo largo de la guía de ondas *B (curva superior)* se puede modular la amplitud del haz óptico de salida *(curva inferior)* de forma tal que su envolvente se convierta en una réplica del haz de impulsos.

Otra aplicación del principio de reflexión de Bragg en guías de ondas onduladas ha sido puesta a punto recientemente en el laboratorio del Instituto de Tecnología de California por mis colegas Chi-Shain Hong, Jeffrey B. Shellan, Abraham Katzir y Alexander C. Livanos. Su diseño consiste en introducir en una guía de ondas transparentes una sección ondulada de período variable. El haz de entrada se introduce en la guía según un ángulo de 45 grados respecto a la dirección de la ondulación. Algunas longitudes de onda que satisfacen la condición de Bragg (que en este caso exige que el período de las ondulaciones sea igual a un múltiplo de la longitud de onda dividido por la raíz cuadrada de 2) se reflejan según una dirección perpendicular a la del haz incidente.

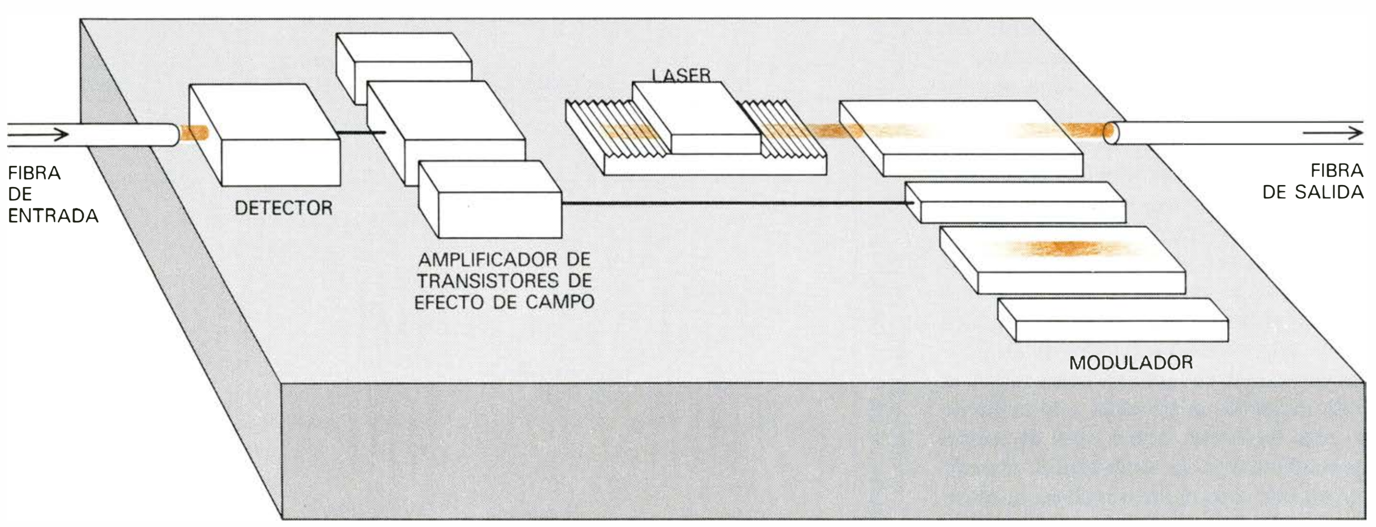
Puesto que el período de las ondulaciones varía a lo largo del recorrido de la luz que se propaga en el interior de la guía, se reflejan longitudes de onda diferentes en lugares diferentes. Este hecho puede dar lugar a determinadas aplicaciones en los casos en que un cierto número de haces luminosos, cada uno de los cuales tiene una longitud de onda diferente y transporta una información diferente, se transmiten a lo largo de un canal único (por ejemplo, una fibra óptica). En la estación receptora, los haces se separarían espacialmente por medio de otro filtro de Bragg con período variable y se enviarían hacia una serie de detectores diferentes con el fin de poder recuperar la información transportada por cada uno de ellos.

Cuando en un láser semiconductor-diodo se incorporan dos secciones onduladas, una a cada lado de la sección de amplificación, dichas secciones desempeñan el mismo papel como reflectores que los espejos terminales de un láser. El alto grado de selectividad de estos reflectores da lugar a que la luz que emite el láser sea extremadamente monocromática (es decir, con una longitud de onda única). En consecuencia, se puede seleccionar la longitud de onda característica de un láser de semiconductor-diodo intercalando en la estructura del láser una sección ondulada cuyo período sea el que corresponda a dicha longitud de onda. Los láseres que operan según este principio fueron puestos en funcionamiento por primera vez en 1975 por Huan-Chun Yen, Michiharu Nakamura y yo mismo en el Instituto de Tecnología de California.

El período de las redes onduladas que se utilizan en las aplicaciones descritas anteriormente es del orden de una fracción de micrometro; este valor corresponde a algunos miles de rayas por milímetro. Estas redes ultrafinas se obtienen mediante una técnica fotolitográfica utilizando un método de exposición holográfica. Dicha técnica se basa en las figuras de interferencia que forman dos rayos de láser en una película fotosensible delgada depositada sobre la superficie a ondular. El período de la red viene determinado por la longitud de onda de los rayos láser y su ángulo de incidencia.

El estudio de la reflexión de la luz bajo la condición de Bragg ha conducido recientemente al descubrimiento por parte del autor y de sus colegas Pochi A. Yeh, del Cal Tech, y Alfred Y. Cho, de los Laboratorios Bell, de una nueva forma de propagación de ondas luminosas en las proximidades de la interfase de un material semiconductor estratificado y un medio homogéneo. Si el período del material estratificado (la distancia de repetición) es aproximadamente igual a la mitad de la longitud de onda de la luz incidente, las ondas luminosas en cuestión no podrán entrar en el seno del material estratificado, puesto que las sucesivas reflexiones procedentes de las interfases próximas estarán exactamente en fase una con otra y, por tanto, se reforzarán. (Esta situación es análoga a la reflexión de Bragg de los rayos X por los planos de un cristal.) Normalmente, la onda incidente, incapaz de entrar en el seno del material estratificado, escaparía al medio homogéneo que rodea a dicho material (aire), pero en este caso será devuelta por reflexión interna. El resultado neto de todo este proceso es un modo de propagación en el que la mayor parte de la energía de la luz queda confinada en las primeras capas del material en cuestión y se propaga en una dirección paralela a los planos definidos por las interfases que separan dichas capas.

La fabricación de estos materiales estratificados de alta precisión ha sido posible gracias a los avances logrados en una técnica de crecimiento de cristales, llamada epitaxia de haces moleculares, que hace posible controlar el espesor de las oportunas capas con una precisión del orden de algunos angstrom. Las guías de ondas ópticas basadas en el principio de la reflexión de Bragg están



CIRCUITO OPTICO COMPLETO, que pueden construirse utilizando la tecnología de que se dispone en la actualidad, sobre un pequeño cristal de material semiconductor. Por ejemplo, el presente circuito se podría utilizar como estación relé en un enlace de comunicación óptica en el que la señal portadora de información degradada y atenuada como consecuencia de su transmisión a grandes distancias a lo largo de una fibra óptica, se detecta, se amplifica y se reforma electrónicamente. A su vez, la señal de voltaje resultante podría servir para modular la salida de un láser de semiconductor-diodo en un dispositivo de acoplamiento direccional del tipo representado en la página anterior. Se está trabajando en circuitos más complejos.

exentas de algunas de las limitaciones propias de las guías de ondas convencionales y se espera que jueguen un importante papel en todas aquellas aplicaciones en las que conviene un elevado nivel de selectividad en las longitudes de onda, así como en los casos en que no se pueden utilizar los sistemas de guías de ondas ordinarios.

El hecho de que la mayoría de las guías de ondas hayan sido construidas con materiales cristalinos abre toda una serie de posibilidades que los circuitos de microondas no son capaces de ofrecer. Por ejemplo, el índice de refracción de muchos materiales cristalinos se puede controlar mediante la aplicación de un campo eléctrico (efecto electroóptico). Puesto que las propiedades de un material dado como guía de ondas dependen fuertemente de su índice de refracción, la aplicación de un campo eléctrico a una guía de ondas transparentes puede modificar drásticamente la naturaleza de la luz propagada. Este método de control electroóptico se puede aplicar fácilmente a la modulación, conmutación y acoplamiento rápidos de haces luminosos.

Un ejemplo de control electroóptico del proceso de transmisión a través de una guía de ondas lo tenemos en el denominado acoplamiento direccional. Este efecto se basa en la utilización de dos guías de ondas paralelas lo suficientemente próximas entre sí como para que la porción más periférica del haz luminoso que se propaga a lo largo de una de las guías de ondas alcance a la otra guía y viceversa [*véase la ilustración superior de la página 45*]. Bajo estas condiciones, la onda que se propaga por una de las guías de ondas *A* se irá escapando gradualmente hacia la segunda guía *B*, siempre que la velocidad de las ondas en ambas guías sea la misma. La igualdad entre las velocidades de las ondas asegura que la excitación de la guía de ondas *B* por el campo procedente de la guía de ondas *A* esté totalmente en fase con la onda que se propaga en dicha guía *B*. Puesto que la potencia total propagada es constante, la excitación de la onda luminosa en la guía *B* debe venir acompañada por una disminución de la potencia en la guía de ondas *A*.

Como resultado de todo este proceso, al cabo de cierta distancia se habrá producido una transferencia total de la potencia desde la guía *A* hasta la guía *B*. Si se aplica un campo eléctrico a lo largo de la guía de ondas *B*, suministrando un cierto voltaje a un par de electrodos próximos, el cambio electroóptico inducido en el correspondiente índice de refracción producirá una cierta alteración en la velocidad de la onda en dicha guía. El desajuste en la velocidad resultante impide la acumulación de potencia en la guía de ondas en cuestión y, como consecuencia, tan sólo se transmite a la guía *B* una cierta fracción de la potencia total de la onda luminosa que se propaga a lo largo de la guía *A*. Más adelante, la potencia acoplada vuelve a la guía *A*. De hecho, se dispone así de un sistema en el que la potencia que entra en una guía dada puede emerger a través del extremo de la otra guía (cuando el voltaje está "desconectado"), o a través del extremo de la guía citada en primer lugar (cuando el voltaje está "conectado").

Esta conmutación eléctrica de luz guiada se puso de manifiesto por vez primera en el año 1975 por James Campbell y Fred A. Blum, de la Texas Instrument Incorporated, y puede utilizarse en la construcción de redes de distribucion óptica en las que un cierto número de fibras ópticas de entrada se conectan entre sí para formar cualquier tipo de combinación que se desee. Este sistema dispone de un determinado número de guías ópticas de salida. Es importante señalar que el citado proceso se puede utilizar también para modular un haz óptico con una señal eléctrica. Para comprender cómo se lleva a cabo esta modulación, vamos a considerar un flujo de datos codificados en un tren de impulsos de voltaje binarios ("on" –conexión– y "off" –desconexión–) aplicados a lo largo de la guía de ondas *B* [*véase la ilustración inferior de la página 45*], y vamos a suponer que un haz continuo de láser penetra por el extremo de la guía de ondas *A*. Durante el tiempo de conexión de un impulso ("on"), la potencia emerge por el extremo de la guía de ondas *A*; durante el tiempo de desconexión del impulso ("off"), la potencia emerge por el extremo de la guía de ondas *B*.

En virtud de lo que antecede, el tren de impulsos eléctricos se reproduce bajo la forma de una modulación de la envolvente del haz portador óptico; en otras palabras, la representación gráfica de la distribución espacial de la intensidad de la onda luminosa que se propaga es análoga a la forma ondulatoria del voltaje responsable de la modulación. La información se puede transportar por medio del haz luminoso que se propaga por el interior de la fibra o a través de la atmósfera. Más tarde, un detector óptico tal como un fotomultiplicador o un semiconductor-diodo puede reconvertir el haz luminoso en una réplicas casi perfecta de la señal eléctrica original.

Ciencia y sociedad

Robert A. Millikan

En su tiempo, Robert A. Millikan (Nobel de física en 1923) fue el más famoso científico norteamericano. Aunque su fama se debe, sobre todo, al experimento de la gota de aceite, con que alcanzó a medir la carga del electrón, a lo largo de su fecunda carrera aportó importantes contribuciones al estudio del efecto fotoeléctrico, espectros de chispas calientes y, en particular, rayos cósmicos.

Nacido en 1868, hijo de un clérigo, y criado en un ambiente rural, Millikan llegó a tomar contacto con la física de la mano de su profesor de griego ("Quien tenga éxito con mis clases de griego puede explicar física", le dijo). Se doctoró en física por Columbia, y tras un viaje de estudios por Europa, donde tuvo por maestros, entre otros, a Max Planck, Walther Nernst y Henri Poincaré, comenzó a ejercer la docencia en la Universidad de Chicago, en 1896. Allí escribió –solo o en colaboración– una serie de libros y manuales de laboratorio. Uno de ellos superó la cifra de millón y medio de ejemplares vendidos, entre 1906 y 1952.

Dándose cuenta de que en Chicago primaba la investigación sobre la docencia, Millikan se dedicó exclusivamente a la primera a partir de 1908. La investigación que le dio más fama, la de la gota de aceite, se inició en Chicago en 1909. Los detalles del experimento son muy conocidos, y pueden verse en cualquier manual. Millikan publicó en 1913 el valor por él obtenido, tras largos trabajos y correcciones de una serie de constantes físicas cuyo valor no se había determinado con la necesaria exactitud para su experimento: $e = 4{,}774 \quad 0{,}009 \times 10^{-10}$ unidades electroestáticas de carga. Este valor permaneció en vigor durante una generación. Pero tan importante como el valor era la clarísima información que implicaba sobre la naturaleza del electrón. Antes de él se conocía la razón carga/masa; pero no se había podido probar que todos los electrones fuesen partículas idénticas. El experimento de Millikan dio en este sentido una prueba concluyente.

Finalizado el trabajo sobre la carga del electrón, Millikan retornó a uno de sus primeros temas: el efecto fotoeléctrico. Hacia 1915 consiguió demostrar que la ecuación de Einstein es válida en todos sus extremos, proporcionando así el mejor valor de la constante de Planck, h, disponible en aquel entonces. No obstante, Millikan no creía haber confirmado con su trabajo la teoría cuántica de la luz, que lo explicaba. Tan grande era el peso de la teoría ondulatoria, que estaba convencido, como otros muchos físicos de su tiempo, que la ecuación tenía que basarse en una hipótesis falsa.

La Primera Guerra Mundial, que para los Estados Unidos comenzó en 1917, interrumpió su actividad investigadora. Era por entonces uno de los más distinguidos físicos norteamericanos, con participación creciente en las materias de su especialidad y en la National Academy of Sciences. Fue además uno de los primeros en establecer lazos entre los laboratorios de investigación y la industria, llegando a consultor de la Western Electric, e introduciendo en la industria a varios de sus discípulos. A principios de 1917, para contribuir a la movilización del esfuerzo científico en favor de la guerra, Millikan fue a Washington, como vicepresidente y director de investigación del recién creado National Research Council. Pronto pasó a formar parte de las Fuerzas Armadas, dirigiendo con el grado de teniente coronel trabajos de meterología, comunicaciones e instrumentación aeronáutica. Mientras, el NSC inició y desarrolló un proyecto para la detección de submarinos.

El éxito de Millikan en la movilización bélica no fue hazaña de poca monta. Para conseguir sus propósitos tenía que actuar con tacto, ser persistente y utilizar juiciosamente sus influencias. Millikan aprendió a "vender ciencia" a diversidad de gentes, tanto civiles como militares. Tanto impresionaron los resultados al astrofísico George Ellery Hale y al físico-químico Arthur A. Noyes, que, a propuesta suya, Millikan fue a residir a Pasadena, para convertirse en cabeza del nuevo y munificamente dotado Cal Tech y director de su laboratorio de física. Reclutó el mejor profesorado disponible, consiguió que los ricos californianos consideraran privilegio ser invitados a donar al instituto y amplió en gran medida las instalaciones. La magnífica gestión de Millikan, la visión de Hale y la prudencia de Noyes convirtieron al Cal Tech en un gran éxito, casi de la noche al día.

En el período de entreguerras, Millikan fue influyente miembro de la National Academy y del National Science Council, que contribuyó a administrar con el fin de mejorar la calidad de la física norteamericana, especialmente en los estudios teóricos. Durante la Segunda Guerra Mundial, Millikan fue traspasando algunas de sus atribuciones en Cal Tech a miembros más jóvenes, y en 1946 dejó su posición como cabeza del ejecutivo, aunque permaneció en el instituto hasta su muerte, en 1953. Sus compromisos administrativos de los años de Cal Tech no le impidieron dar cursos de física atómica.

Terminados los trabajos sobre carga del electrón y efecto fotoeléctrico, que le valieron en 1923 el premio Nobel, Millikan retornó a investigaciones interrumpidas a causa de la guerra: los espectros de chispas calientes y la "radiación penetrante", que más tarde bautizaría como rayos cósmicos.

En 1920, ya en Cal Tech, Millikan se embarcó, con Ira S. Bowen, en el estudio sistemático del espectro ultravioleta de los elementos más ligeros. Los trabajos de ambos contribuyeron a cerrar el intervalo entre las frecuencias ópticas y los rayos X. En su trabajo descubrieron que al arrancar de los átomos los electrones de valencia se producían espectros con pares de líneas muy próximas (dobletes), descubrimiento que contribuyó no poco a postular el spin del electrón en 1925.

Mientras realizaba estos estudios espectrales, Millikan reanudó su investigación sobre radiación penetrante, descubierta al observar que la ionización atmosférica aumenta con la altitud, lo que sugería que algún tipo de radiación debería llegar desde el cosmos. No obstante, muchos científicos opinaban que la ionización se debería a causas terrestres, como radiactividad. Para zanjar la cuestión, Millikan se propuso en el verano de 1925 medir la variación de ionización con la profundidad en los lagos Muir y Arrowhead, en las montañas de California, cuya diferencia de altitud sobre el mar es de unos 2000 metros. Las mediciones electroscópicas probaron que la intensidad de ionización a una profundidad cualquiera del lago Arrowhead (el más bajo) era idéntica a la correspondiente a una profundidad 1,80 metros mayor en el Muir. Como la capa atmosférica que separa ambas superficies tiene la misma capacidad de absorción que 1,80 metros de agua, los resultados probaron positivamente que la radiación provenía del cosmos.

Siendo mucho más penetrantes que los más fuertes rayos gamma entonces conocidos, Millikan consideraba que no podía tratarse de partículas cargadas eléctricamente. Millikan aventuraba la conjetura de que se tratase de fotones, cuantos de radiación electromagnética. Las trayectorias de los fotones no sufrirían desviaciones provocadas por el campo magnético terrestre; lo contrario le sucedería a partículas eléctricamente cargadas, con lo que se recibiría mayor radiación en las latitudes extremas que en las cercanas al ecuador. Millikan realizó experimentos en

América del Sur y no detectó variaciones de intensidad con la latitud, pero otros investigadores, particularmente el holandés Jacob Clay, en 1927, sí comenzaban a detectarlas. En 1929 dos físicos alemanes (Bothe y Kohlhorster) encontraron pruebas de que al menos una importante fracción de los rayos cósmicos estaba formada por partículas cargadas, de enorme energía. Hubo confirmación por otros conductos, y finalmente, en 1932, se demostró incontrovertiblemente que sí existía "efecto de latitud" (Compton).

Millikan impugnó acaloradamente las afirmaciones de que los rayos cósmicos fueran partículas cargadas. Tras experimentos realizados por colaboradores suyos, Millikan tuvo que admitir, en 1933, que al menos parte de los rayos cósmicos tenía que estar formada por partículas cargadas, pero insistió en que tales rayos eran "secundarios", producidos por los encuentros entre "primarios" y núcleos de átomos de la atmósfera. Tenazmente, persistió en la creencia de que los rayos primarios, o al menos parte de ellos, eran fotones, durante los 20 años siguientes, a pesar de abrumadora evidencia contraria y de la opinión de casi todos los especialistas.

¿Por qué este empecinamiento de Millikan? Hay varias posibles explicaciones. Tenía 65 años, y se había hecho egocéntrico y dogmático. Se había declarado, pública y notoriamente, partidario de la teoría fotónica de los rayos cósmicos. Posiblemente influyeran también sus convicciones religiosas. Miembro de la secta universalista, y próximo a creencias deístas, Millikan veía la mano del Creador en el origen de los rayos cósmicos (o al menos en su particular teoría), siendo éstos fotones: serían generados cuando cuatro átomos de nitrógeno se condensasen en uno de helio, cuando 14 hidrógenos dieran un nitrógeno, o 16, un átomo de oxígeno. Los correspondientes defectos de masa serían radiados en forma de energía electromagnética. Los rayos cósmicos serían "los llantos de átomos recién nacidos", frase de Millikan que alcanzó gran popularidad. Más tarde Millikan propuso otras teorías, pues el problema nunca dejó de interesarle, realizando una larga serie de observaciones a todas las altitudes y profundidades. Incluso la teoría del nacimiento de átomos proporcionó dividendos: deseoso de obtener medidas directas de las energías de los rayos cósmicos, hizo que el joven Carl D. Anderson estudiara el comportamiento de los rayos cósmicos en una cámara de niebla, lo que le llevó a la detección del positrón en 1932.

Millikan pudo haber estado orgulloso de los últimos años de relativo fracaso. Las preguntas que suscitó sobre la política de investigación no han perdido ninguna vigencia. Y aunque equivocado sobre la naturaleza y origen de los rayos cósmicos, el problema no es de poca monta. En 1950 escribía: "Nuevas pruebas han aparecido, contrarias a esta hipótesis (la aniquilación atómica), pero los datos experimentales que contiene (su artículo) son válidos. El verdadero origen de los rayos cósmicos sigue siendo un misterio". El misterio aún no ha sido resuelto.

Primos que progresan

Para números enteros muy grandes resulta muy difícil determinar si se trata de números compuestos o de números primos, es decir, si admiten o no divisores distintos del propio número y de la unidad. Aunque para probar que un número es primo basta comprobar que no es divisible por ninguno de los números primos anteriores a él, al crecer el número a estudiar la puesta en práctica de tal procedimiento exige tanto tiempo que resulta prohibitivo aun disponiendo de los más potentes ordenadores. Hay, no obstante, ciertos tipos de números para los que han podido encontrarse criterios algorítmicos mucho más eficientes con los que comprobar si son o no son números primos. Por ejemplo, en 1971 Bryant Tuckerman, del Thomas J. Watson Research Center, de la International Business Machines Corporation, consiguió probar que el número $2^{19.937}-1$ es primo. Este número, que escrito en la notación ordinaria de base 10 tiene 6 x 002 cifras, era mayor que cualquier otro de los primos entonces conocidos. En octubre del pasado año, Laura Nickel y Curt Noll demostraron que $2^{21.701}-1$ (número que tiene 6533 cifras) también es primo.

Los números primos expresables en la forma 2^p-1, donde p es un entero primo, se llaman números primos de Mersenne, en honor de un matemático del siglo XVII que los estudió. A lo largo del siglo pasado casi todos los números primos descubiertos fueron primos de Mersenne, pues existe un algoritmo eficiente para comprobar si un número de la forma 2^p-1 es primo o compuesto. (Dado que 2^n-1 no puede ser primo sin que lo sea el exponente n, no es preciso estudiar números de esta forma que tengan exponentes compuestos.) Tal algoritmo, llamado criterio de Lucas-Lehmer, fue creado por el matemático francés Édouard Lucas en 1876, y más tarde perfeccionado por D. H. Lehmer en 1930. Para comprobar el carácter primo de números de esta forma se genera una sucesión de números, comenzando en 4, en la cual cada término (u_i) de la sucesión es igual al cuadrado del término precedente, menos 2 (u_{i-1}^2-2), reducido módulo 2^p-1. Si el p-ésimo término de la sucesión es igual a 0 entonces 2^p-1 es primo; en caso contrario, 2^p-1 es número compuesto.

El número de etapas de computación precisas para ejecutar tal procedimiento directamente es proporcional al cubo del exponente p. El número de pasos necesarios para probar que un número es primo a base de dividirlo por los primos que lo preceden es proporcional a la raíz cuadrada del número dado, pues al dividir por los primos menores que la raíz cuadrada del número quedan de manifiesto todos sus divisores primos. Al aumentar p, la tercera potencia de p crece mucho más lentamente que la raíz cuadrada de 2^p-1. Más aún, incluso para valores no muy grandes de p, el criterio de Lucas-Lehmer es significativamente más eficiente que el otro.

Al aproximarse a 20.000 el exponente p, la aplicación del criterio de Lucas-Lehmer plantea un complejo problema de programación, que exige multiplicar y restar números de miles de dígitos. (Conviene notar, sin embargo, que en notación binaria el número 2^p-1 se escribe como sucesión de p signos 1.) Nickel y Noll, trabajando en el centro de cálculo de la filial de la Universidad del Estado de California en Hayward, escribieron hasta cinco versiones diferentes de su programa antes de comenzar a ensayar números, tarea que emprendieron el pasado mes de septiembre. Tuckerman había ensayado todos los exponentes primos hasta 21.000; por ello Nickel y Noll comenzaron a partir de ahí. Tras unas 350 horas de trabajo de computador, el programa consiguió reconocer que 21.701 era exponente de uno de los números primos de Mersenne.

El descubrimiento de este nuevo número primo ofrece particular interés. Ante todo, como no ha podido probarse que el conjunto de números primos de Mersenne sea infinito, el descubrimiento de uno nuevo tiene siempre interés. Se ha observado además que en ocasiones los números primos de Mersenne se presentan en pares relativamente cercanos, conocidos por "pares gemelos", como por ejemplo $2^{2203}-1$ y $2^{2281}-1$. El nuevo número primo recién hallado –máximo de los conocidos– se encuentra francamente cerca del anterior. Si se tratase de un par gemelo, es muy posible que el próximo primo de Mersenne se encuentre en la frontera de nuestra capacidad de cómputo. Por otra parte, es posible que estos dos números no sean gemelos, y en tal caso $2^{21.701}-1$ es posible que tenga un hermano gemelo mayor no muy lejano. Para esclarecer tales cuestiones, Nickel y Noll están actualmente revisando su programa, poniéndolo a punto para ensayar números de exponente todavía mayor, con la esperanza de encontrar otro número primo más alto aún.

La estructura profunda de los continentes

Las partes más antiguas de los continentes parecen tener zonas de raíces profundas que viajan con ellos cuando las placas se desplazan. Estas zonas pueden ser químicamente distintas de las rocas que las rodean

Thomas H. Jordan

A lo largo de las dos últimas décadas, los científicos que estudian la tierra han ido reuniendo innumerables datos geológicos, geofísicos y geoquímicos que han permitido desarrollar una perspectiva nueva sobre la historia, la estructura y el comportamiento dinámico de las envolturas más externas del planeta. Ello ha sido posible gracias a la elaboración y sistematización de la teoría de la tectónica de placas. La teoría se ha mostrado eficaz, sobre todo, a la hora de explicar cómo la corteza oceánica basáltica, cuyo espesor es de unos siete kilómetros, se produce continuamente en las crestas de las dorsales centro-oceánicas debido al ascenso de magma fundido procedente del manto subyacente, cómo se mueve por la superficie (a velocidades del orden de centímetros por año, formando la capa superior de enormes placas rígidas con un espesor diez veces mayor) y cómo, esa misma corteza oceánica, es consumida de nuevo en el manto en las zonas de subducción adyacentes a las fosas oceánicas profundas. Esta teoría ha tenido menos éxito al explicar la formación y estructura de los continentes.

La corteza continental, con un espesor medio de 35 kilómetros, es más ligera que la corteza oceánica y más rica en silicio y potasio. En la teoría de la tectónica de placas, se considera que la corteza continental flota y es el producto resultante de la fusión y refusión que, por acreción, la ha ido formando durante los ciclos que conformaron los fondos oceánicos a lo largo de grandes períodos de tiempo. El movimiento de los continentes a través de la superficie terrestre, o deriva continental, muy estudiado, por otra parte, se ha explicado como si se tratara del simple arrastre pasivo de esta corteza ligera al formar parte de una placa en movimiento. Por su carácter flotante, la corteza continental no puede mezclarse de nuevo con el manto, por lo menos en grandes cantidades, a través de procesos tectónicos como la subducción; tiende, por contra, a flotar sobre el manto más denso como la escoria sobre el hierro fundido. Si prescindimos de la anterior distinción, la tectónica de placas no postula diferencias esenciales entre los movimientos de los continentes y los del fondo oceánico, o entre las estructuras subcorticales de las placas continentales y las de las placas oceánicas.

Pero he aquí que los datos sismológicos revelan profundos contrastes entre las estructuras oceánicas y las continentales que se extienden muy por debajo de la base de la corteza. Bajo las partes más antiguas de los continentes, existen, al parecer, zonas de raíces profundas, de varios centenares de kilómetros de espesor, que viajan junto con los continentes cuando las placas se mueven. El descubrimiento de estas zonas de raíces ha puesto en entredicho alguno de los principios básicos de la teoría de la tectónica de placas. El esquema resultante de la estructura profunda de los continentes proporciona nuevos puntos de vista relativos a los procesos mecánicos y físicos que controlan la evolución y la tectónica continentales.

Estabilidad cratónica

Las rocas más antiguas de la corteza continental se encuentran en el complejo que constituye el zócalo de los escudos y plataformas continentales. Los escudos son extensas superficies levantadas, desprovistas esencialmente de sedimentos recientes. Las plataformas son depresiones, amplias y de poca profundidad, que afectan al complejo del zócalo, y que están rellenas de rocas sedimentarias casi horizontales. Las plataformas y los escudos forman los cratones, bloques estables que representan los núcleos de las masas continentales actuales. Si bien la mayor parte de las rocas del zócalo que afloran en los escudos o que se hallan bajo los sedimentos de las plataformas han sufrido procesos de metamorfismo durante episodios orogénicos antiguos (formación de cordilleras de montañas), han permanecido inalteradas durante períodos de tiempo muy largos, del orden de mil millones de años, o más.

Sin embargo, los continentes contienen algo más que sus núcleos estables cratónicos. En contraste con los cratones, existen las zonas orogénicas actuales de los continentes: enormes áreas sometidas a profundas deformaciones en tiempos relativamente recientes debido a la actividad tectónica resultante de la convergencia de dos placas con movimientos opuestos. En una zona orogénica activa, la corteza no se comporta de modo uniforme; y la tectónica de placas no alcanza a describir sus movimientos. El contraste entre cratones y zonas orogénicas puede observarse hoy en día en Asia meridional, donde se produce una deformación causada por colisión de los cratones arábigo e indio (que se mueven hacia el norte) con el continente euroasiático. Esta violenta colisión "ha arrugado" Asia, produciendo una gran cordillera de montañas que se extiende desde la península de Anatolia hasta los mares que bordean el océano Pacífico, pasando a través de Oriente Medio, el sur de la Unión Soviética y la mayor parte de China [véase "La colisión entre la India y Eurasia", por Peter Molnar y Paul Tapponnier; INVESTIGACION Y CIENCIA, junio, 1977].

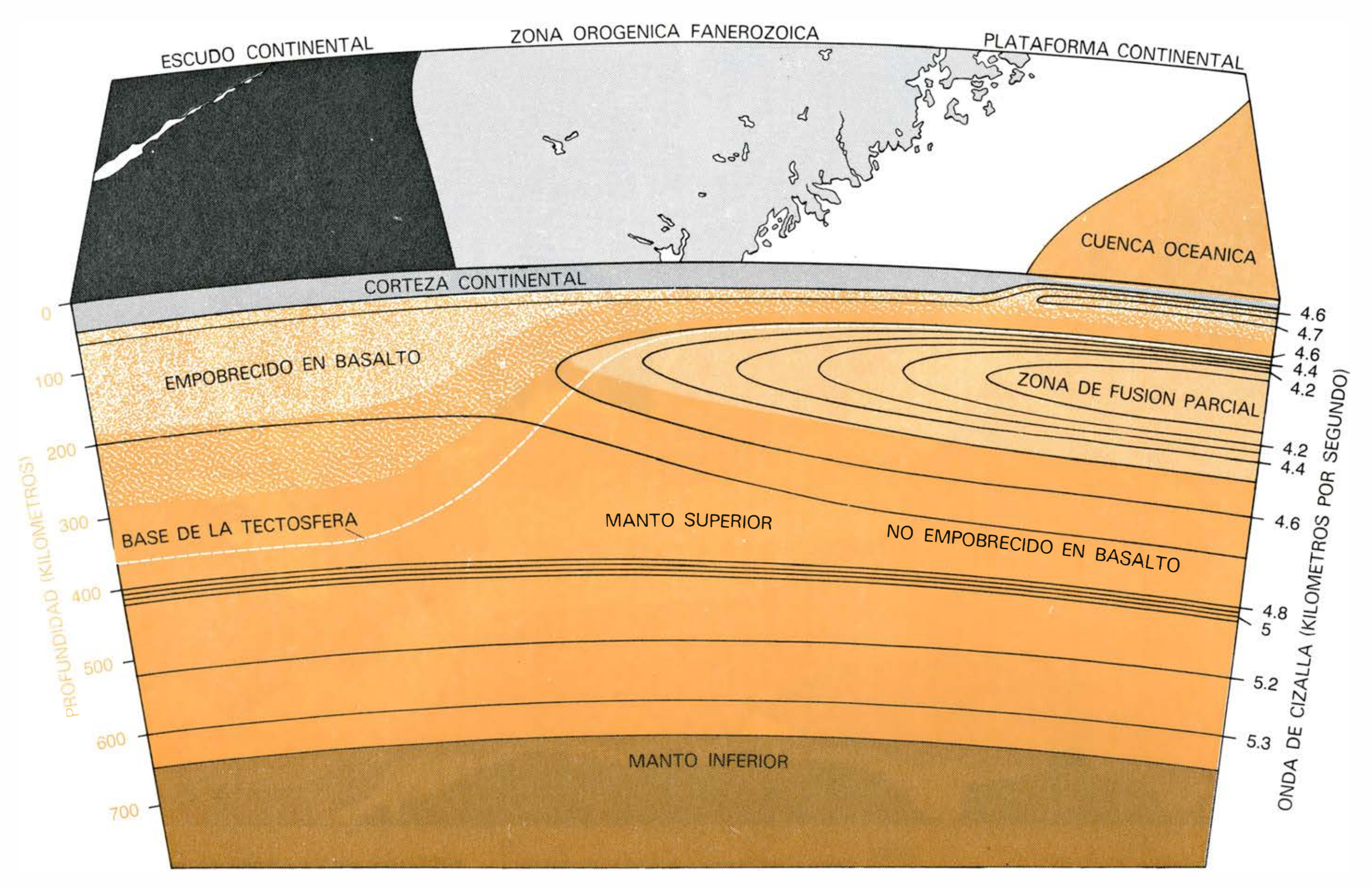
En este acontecimiento tectónico que se está produciendo en la actualidad hay que resaltar que el cratón de la India continúa moviéndose frontalmente hacia Asia, a una velocidad de cinco centí-

metros por año; ello provoca el cabalgamiento y levantamiento de elevadas montañas en una extensión de 10 millones de kilómetros cuadrados, sin que se produzca una deformación en el propio cratón. Llegar a comprender la dinámica de este comportamiento tectónico misterioso, al objeto de explicar la violenta deformación que se produce en Asia y la persistencia de un subcontinente índico no deformado, constituye un reto muy sugestivo que deberá afrontar la ciencia de la geofísica. Cualquier modelo válido de estructura continental y comportamiento tectónico ha de tener en cuenta este contraste, debiendo basarse en información procedente de las zonas profundas del manto en donde se originan las fuerzas que han configurado los continentes.

El comportamiento tectónico de los continentes está controlado, sobre todo, por la estructura mecánica de la litosfera, que se define funcionalmente como la envoltura resistente y más externa de la tierra y que está compuesta por la corteza y la parte superior del manto. A ella hay que atribuirle la rigidez de las placas. Los estudios sobre el campo gravitatorio terrestre y sobre la respuesta de la tierra a la carga y descarga rápidas de su superficie, debida a grandes masas, como pueden ser los hielos continentales y las cordilleras, sugieren que la litosfera posee un espesor medio efectivo de unos 100 kilómetros. Bajo ella se encuentra la astenosfera, o envoltura blanda, que constituye una región del manto donde cualquier tensión, por pequeña que sea, determina que los materiales componentes se comporten como un fluido. De acuerdo con el modelo de la teoría de la tectónica de placas, la litosfera forma las placas y éstas se mueven con respecto a la astenosfera. Dentro de esta última, se concentran las deformaciones de cizalla asociadas al movimiento de las mismas.

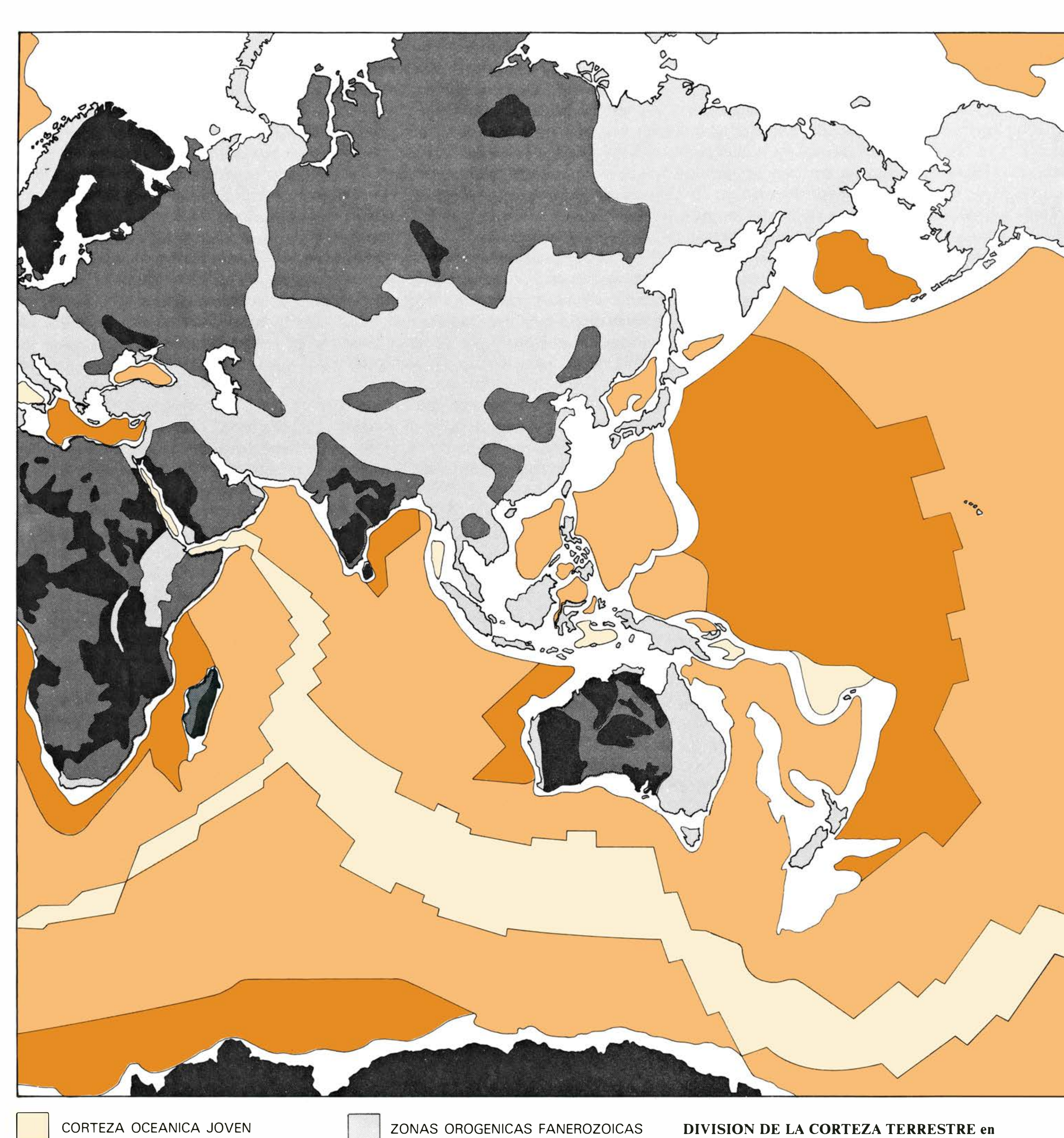
La resistencia de la litosfera deriva, en gran parte, de la envoltura superior del manto; la resistencia del material que compone a este último varía con la temperatura. Las mediciones de gradientes térmicos realizadas cerca de la superficie demuestran que las temperaturas registradas en el interior de la litosfera aumentan rápidamente con la profundidad; la tasa de este aumento varía mucho con la situación geográfica. Los estudios de laboratorio y los datos de campo indican que, a determinadas temperaturas, las rocas corticales, cuya composición varía desde los basaltos hasta los granitos, son significativamente menos resistentes que las peridotitas, principal constituyente del manto. Debido al poco espesor de la corteza oceánica, las peridotitas del manto se encuentran a poca profundidad, donde las temperaturas son bajas. Según parece, este horizonte de peridotita fría, cercano a la superficie, da a la litosfera oceánica la mayor parte de su resistencia frente a la deformación tectónica. Por contra, la corteza continental es gruesa y, consiguientemente, las temperaturas existentes en su base son más elevadas que las de la base de la corteza oceánica. Bajo los continentes, la parte superior del manto es más débil que su homóloga de las zonas oceánicas; esta debilidad parece explicar, al menos en parte, por qué las zonas orogénicas se deforman con tanta facilidad.

En este caso, ¿por qué los cratones se han mantenido estables y sin deformar a lo largo de períodos de tiempo tan prolongados de la historia geológica? La



MODELO HIPOTETICO de la zona de transición bajo un margen continental típico. En él se observan algunos rasgos que podrían caracterizar el manto superior de la tierra en un corte que se extendiera desde el escudo canadiense (*izquierda*) hasta la cuenca del océano Atlántico (*derecha*). Las grandes variaciones laterales de las velocidades sísmicas de la onda de cizalla observadas en medidas recientes se indican mediante líneas negras. Según la hipótesis del empobrecimiento basáltico sostenida por el autor, el manto superior de la tierra se compone de una asociación mineral denominada peridotita (*color intenso*); ésta, bajo los continentes, se halla empobrecida en determinados elementos basálticos. La densidad del punteado en blanco indica el grado de empobrecimiento. La línea blanca de trazos representa la profundidad mínima aproximada de la base de la tectosfera, término éste adoptado por el autor para definir el volumen de las placas tectónicas. La zona de fusión parcial se encuentra bajo la corteza oceánica.

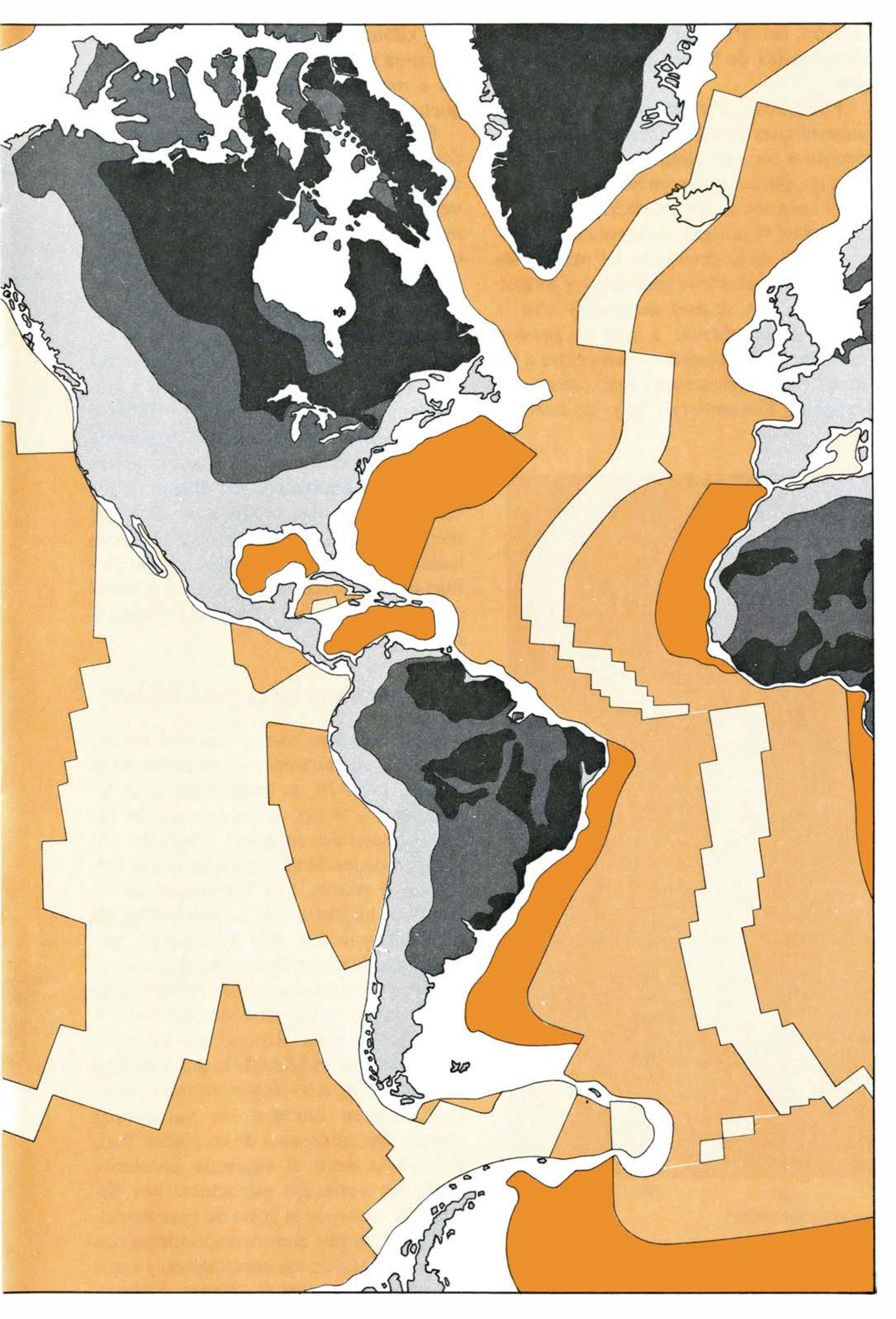
respuesta se halla en estrecha relación con el problema de la estructura continental profunda. Se puede aducir que, si bien las temperaturas en la base de la corteza continental exceden generalmente a las de la base de la corteza oceánica, el gradiente térmico es casi siempre inferior (la temperatura aumenta de un modo más lento con la profundidad) en el manto cratónico: las temperaturas necesarias para que la astenosfera fluya no se alcanzan bajo los cratones hasta profundidades mayores que las correspondientes a las zonas situadas bajo los océanos. Así pues, la temperatura integrada del manto superior y, por tanto, el espesor efectivo de la litosfera, podría ser mayor bajo los cratones que bajo los océanos o las zonas orogénicas. Los geofísicos han supuesto que el aumento de espesor litosférico producido por la exis-



DIVISION DE LA CORTEZA TERRESTRE en siete tipos de rocas principales, según el sistema tectónico de clasificación adoptado en el mapa de estas dos páginas. En general, las zonas coloreadas y las blancas se localizan bajo el nivel del mar; las grises corresponden a la tierra firme (*véanse los signos convencionales de la parte inferior izquierda*). El color más claro se utiliza para señalar las regiones con corteza oceánica de edad

tencia de temperaturas más bajas a profundidades más elevadas explica la estabilidad cratónica.

Más tarde se verá que la cuestión no es tan sencilla como parece. Investigaciones recientes indican que, además de los contrastes térmicos, hay diferencias químicas en el interior de las raíces cratónicas del manto superior que poseen su importancia en la regulación de la tectónica continental. La comprensión del nuevo y aún controvertido modelo del manto superior continental, que incorpora estas diferencias de composición, requiere un examen detallado de la estructura profunda de los continentes.



inferior a los 25 millones de años; el intermedio representa regiones cuya corteza oceánica tiene una edad comprendida entre los 25 y 100 millones de años, y el color más fuerte corresponde a las que tienen una antigüedad superior a los 100 millones de años. Las zonas en blanco son regiones con corteza continental de transición o sumergida, incluidos los márgenes continentales, los arcos de islas y las plataformas marinas. El gris más claro representa regiones cuya corteza continental se ha visto afectada por la actividad orogénica (formación de cordilleras) durante el Fanerozoico, es decir, en el transcurso de los últimos 600 millones de años. El gris intermedio señala plataformas continentales estables con historial sedimentario también fanerozoico, y el más obscuro corresponde a los escudos continentales y a las plataformas en donde no se han depositado sedimentos desde el Precámbrico.

La estructura elástica de la tierra

La mayoría de los conocimientos adquiridos sobre la estructura del interior de la tierra se deben a la sismología. Cuando fuentes potentes de ondas elásticas (terremotos y grandes explosiones) "iluminan" el interior, las imágenes registradas por los sismógrafos indican variaciones especiales de las propiedades elásticas del globo. La respuesta elástica a una fuente sísmica queda adecuadamente descrita por tres parámetros: la densidad, la velocidad de las ondas de compresión (cuyo movimiento está polarizado paralelamente a la trayectoria de la onda) y la velocidad de las ondas de cizalla (cuyo movimiento está polarizado transversalmente a la trayectoria). Cada parámetro elástico es una función de la presión, la temperatura y la composición química.

Desde el punto de vista sismológico, se puede suponer casi siempre que la tierra tiene una estructura esférica simétrica cuya presión, temperatura y composición (y, por tanto, sus parámetros elásticos) varían solamente en función de su distancia al centro de la misma. Este es, por ejemplo, el supuesto que se utiliza en los algoritmos que se usan en los computadores para localizar grandes terremotos, ajustando los tiempos observados de llegada de las ondas sísmicas. Una vez localizado un terremoto en el espacio y en el tiempo, puede computarse el intervalo que tarda una onda sísmica en llegar a una estación cualquiera. Si la tierra fuera realmente simétrica y esférica, es decir, si sus propiedades elásticas variaran sólo en función del radio, los tiempos que las ondas tardan en llegar dependerían únicamente de la distancia, medida sobre un círculo máximo, que separa la fuente y la estación receptora, pero no dependerían de las coordenadas específicas de estas últimas. Sin embargo, en la práctica se observan pequeñas, aunque significativas, variaciones geográficas en los tiempos empleados en llegar a diferentes estaciones, con magnitudes características de menos del 1 por ciento de la duración total. El análisis de estas pequeñas diferencias regionales indica que las estructuras asociadas a los cratones persisten en profundidad, como mínimo, hasta varios centenares de kilómetros, es decir, a profundidades de un orden de magnitud mayor que el espesor de la corteza continental.

El estudio sismológico de la heterogeneidad continente-océano viene dificultado por la falta de estaciones de registro

sísmico en las cuencas oceánicas. Para investigar el manto superior en regiones que carecen de estaciones sísmicas, los sismólogos se valen generalmente de las ondas superficiales cuya energía queda confinada cerca de la superficie, y cuyos tiempos empleados para alcanzar los observatorios dependen de la corteza y del manto superior a lo largo de todo el recorrido entre la fuente y la estación receptora. Las velocidades de propagación de los componentes de baja frecuencia de las ondas de superficie son más sensibles a la estructura profunda que los de frecuencias más elevadas; a partir de las mediciones realizadas de esta dependencia de frecuencia, se pueden estimar las variaciones de las propiedades elásticas del manto superior, tanto para las trayectorias continentales como para las oceánicas. Los datos sobre ondas superficiales recogidos hacia 1920 por dos pioneros de la sismología, Beno Gutenberg y Robert Stoneley, demostraron que la corteza existente bajo los océanos es mucho más delgada que la que se encuentra bajo los continentes. De todas formas, los estudios sistemáticos de las variaciones laterales a gran escala existentes bajo la corteza no empezaron hasta finales de la década 1950-60 y principios de la 1960-70, cuando se instalaron las primeras redes globales de sismógrafos de banda ancha normalizados.

Los datos suministrados por las ondas superficiales de estas redes convencieron pronto a los sismólogos de que las diferencias estructurales entre los continentes y océanos deben extenderse muy por debajo de la corteza. A través de la mayor parte de la corteza y del manto, los parámetros elásticos (densidad y velocidades de las ondas) aumentan con la profundidad, debido a que la presión también se incrementa y comprime a las rocas dando estructuras más compactas y rígidas. Sin embargo, bajo las cuencas oceánicas y la mayor parte de zonas orogénicas más activas, se encontró que la velocidad de la onda de cizalla disminuía bruscamente con la profundidad en una región de transición que empieza a unos 50 o 100 kilómetros por debajo la superficie y forma una "zona de baja velocidad", cuyo espesor aproximado es de 100 kilómetros. Bajo los cratones, falta esa zona de baja velocidad, o bien se halla a más profundidad y es menos importante.

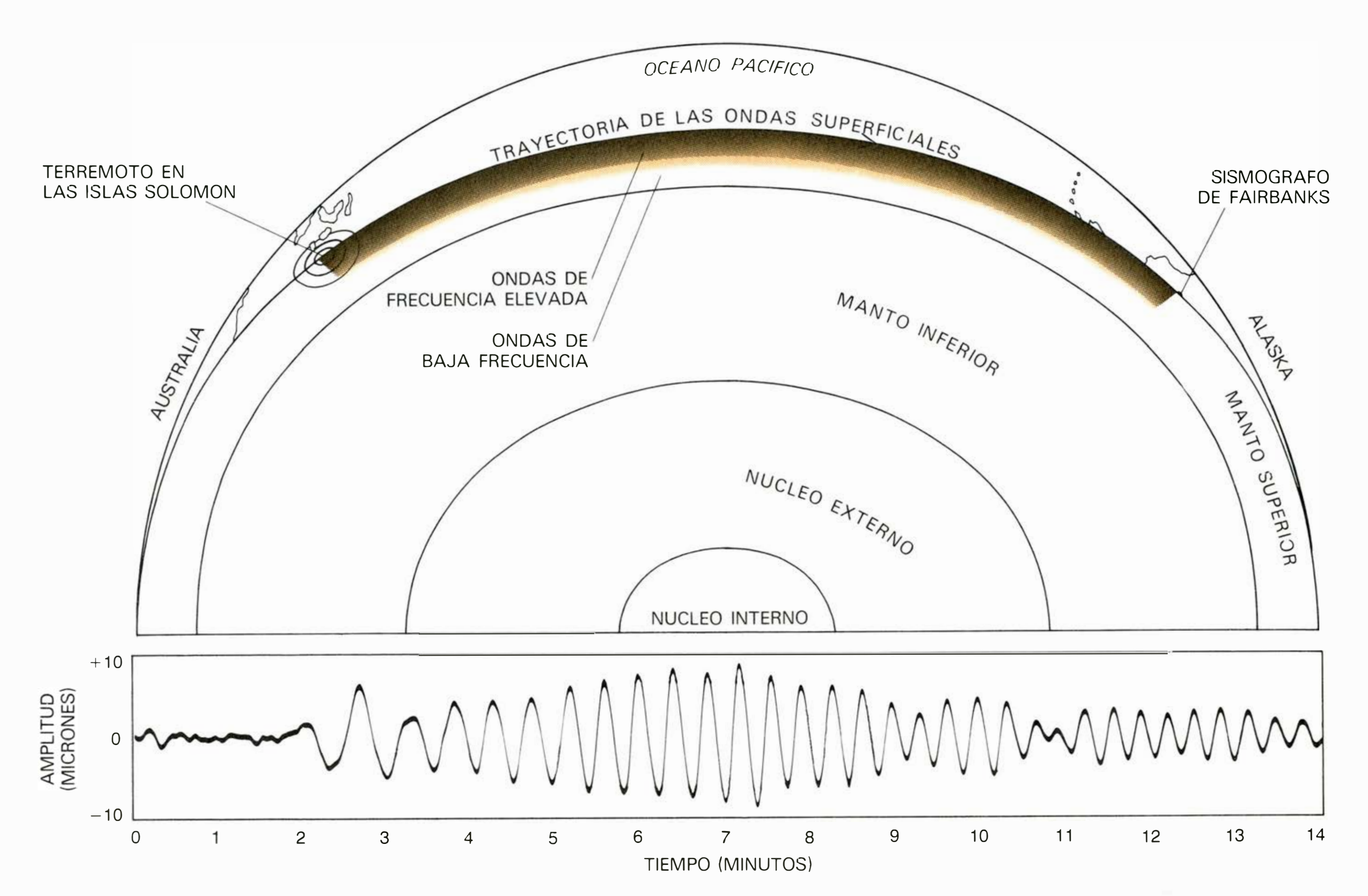
Estos hallazgos sismológicos concordaban bien con los modelos aceptados de la estructura térmica del manto. Las velocidades de las ondas varían generalmente de forma inversa a la temperatura: una pequeña fusión ocasiona un descenso muy importante de la velocidad. Las investigaciones de laboratorio y una serie de consideraciones teóricas indicaron que la zona de baja velocidad del manto superior se debía quizás a una pequeña fusión (1 por ciento, o menos) de las peridotitas del manto, producida por las temperaturas elevadas existentes a unos 100 kilómetros por debajo de los océanos y zonas orogénicas. Por ser menos bruscos los gradientes térmicos localizados bajo los cratones, las peridotitas de esas zonas sólo funden a profundidades más elevadas, si es que llegan a hacerlo.

El problema de la profundidad

Este esquema estructural del manto superior, que apareció a principios de la década 1960-70, se mostró de gran fecundidad a la luz de las teorías, en rápido desarrollo entonces, sobre la expansión de los fondos oceánicos y la tectónica de placas. Los sismólogos tardaron poco en identificar la astenosfera, de fácil deformación, con el material parcialmente fundido de la zona de baja velocidad y, asimismo, la litosfera rígida con el material más frío situado por encima de esa zona. Resultó que al cartografiar la profundidad de la zona de baja velocidad, los sismólogos estaban cartografiando en realidad las variaciones geográficas del espesor de las placas. Pero el modelo llevó al siguiente corolario: cualquier variación estructural por debajo del techo de la zona de baja velocidad no tiene por qué corresponderse con las posiciones de los continentes, ya que los movimientos de las placas deberían estar continuamente readaptando sus posiciones relativas mediante el desplazamiento de material litosférico en relación con la astenosfera.

Desde esta óptica eran desconcertantes los datos sísmicos que demostraban la existencia de velocidades más eleva-

TEMPERATURA de la corteza y del manto superior en función de la profundidad. Las temperaturas de la corteza y la del manto superior varían sistemáticamente en consonancia con la clasificación tectónica de la roca cortical. El aumento más rápido de temperatura con la profundidad se produce bajo las regiones oceánicas jóvenes (*curva 1*) y el menos rápido bajo los escudos continentales precámbricos (*curva 4*). Estas variaciones de la relación temperatura-profundidad parecen ser las responsables de las diferencias geográficas observadas en las velocidades de las ondas sísmicas. Bajo la mayor parte de regiones tectónicas, las curvas de temperatura aumentan de forma lo suficientemente rápida como para cortar el umbral en el que el manto empieza a fundir (*curva blanca*) haciendo que la elasticidad de este último disminuya bruscamente a una profundidad determinada, creando así una zona de velocidad sísmica baja. Sin embargo, bajo los escudos continentales, la curva de temperatura aumenta de un modo más gradual, pudiendo ser que no se produzca la fusión. Esta conclusión no es incompatible con el punto de vista que considera posible la ausencia de la zona de baja velocidad sísmica bajo los escudos.



ONDAS SISMICAS SUPERFICIALES, u ondas cuya energía queda absorbida cerca de la superficie terrestre. Proporcionan a los sismógrafos gran cantidad de información sobre las variaciones de la estructura elástica del manto superior. Así, por ejemplo, el dibujo ideal de esta figura muestra la trayectoria de un tren de ondas superficiales registrado en el observatorio de Fairbanks (Alaska) después del terremoto ocurrido en las islas Solomon en junio de 1970. La parte del sismograma real registrado que se reproduce en la parte inferior muestra una traza característica hecha por una onda de Rayleigh, onda sísmica cuyo movimiento es elíptico en el plano de propagación (aquí sólo se representa la componente vertical del movimiento de la onda). Estas ondas son dispersivas: sus velocidades de propagación dependen de su frecuencia. Las componentes de más baja frecuencia (*en color*) son más sensibles a las variaciones estructurales o profundidades más elevadas del manto que las componentes de frecuencia más alta (*en gris*).

das de las ondas de cizalla bajo los cratones a profundidades superiores a los 200 kilómetros, es decir, por debajo de la litosfera y la zona oceánica de baja velocidad. En 1963, Gordon J. F. MacDonald, de la Universidad de California en Los Angeles, volvió a plantear el problema de la estructura continental profunda. Sus modelos, basados en datos de ondas de superficie y en información sobre el estado térmico del manto superior, implicaban que las temperaturas más bajas y las velocidades de onda más elevadas, características de los cratones en zonas poco profundas, debían extenderse a profundidades del orden de los 500 kilómetros. MacDonald estaba convencido de que una estructura de este tipo resultaba incompatible con la idea de un manto móvil y en convección; concluyó, por tanto, que los movimientos relativamente rápidos de los continentes, necesarios para que se produzca expansión del fondo oceánico y deriva continental, eran improbables, si no imposibles.

Sin embargo, los argumentos de MacDonald contra la deriva no pudieron resistir la avalancha de nueva información que invadió las ciencias de la tierra durante la siguiente década y que convenció a todo el mundo, menos a los más escépticos, de que los movimientos de las placas eran, sin lugar a dudas, reales. Mientras tanto, los datos en los que MacDonald había basado su tesis fueron interpretados de nuevo por los defensores de esta "nueva tectónica global". En 1970, J. G. Sclater y Jean Francheteau, en la Scripps Institution of Oceanography, presentaron modelos térmicos de la litosfera continental y de la oceánica compatibles con la hipótesis de la tectónica de placas. Señalaron que las diferencias entre el perfil térmico continental y el oceánico podrían estar confinados por encima de los 200 kilómetros de profundidad (en vez de extenderse hasta 500 kilómetros, o más, tal como MacDonald había propuesto) y presentaron un modelo con una litosfera oceánica de un espesor aproximado de 100 kilómetros y una litosfera continental de 200 kilómetros. Su modelo fue aceptado por Adam M. Dziewonski, de la Universidad de Harvard, en 1971. Este demostró que, si bien los datos de las ondas superficiales permitían realmente suponer diferencias estructurales entre continentes y océanos, a las grandes profundidades que MacDonald y otros proponían, los datos de que se disponía no obligaban a aceptar este tipo de variaciones. De hecho, los datos podían ser válidos en modelos donde los contrastes estructurales significativos se concentraran por encima de los 200 kilómetros, valor compatible con las conclusiones de Sclater y Francheteau y con la idea general de las placas litosféricas. Las pruebas sobre la existencia de las raíces continentales profundas de MacDonald parecían haberse esfumado.

Sin embargo, algunos sismólogos no quedaron satisfechos con este esquema de la estructura del manto superior. I. Selwyn Sacks, de la Carnegie Institution de Washington, analizó datos procedentes de sismos muy profundos de la costa occidental de Sudamérica y sacó la conclusión de que el espesor de la placa que constituye el cratón sudamericano debe sobrepasar los 300 kilómetros. A un resultado similar llegó Shelton Alexander, de la Universidad estatal de Pennsylva-

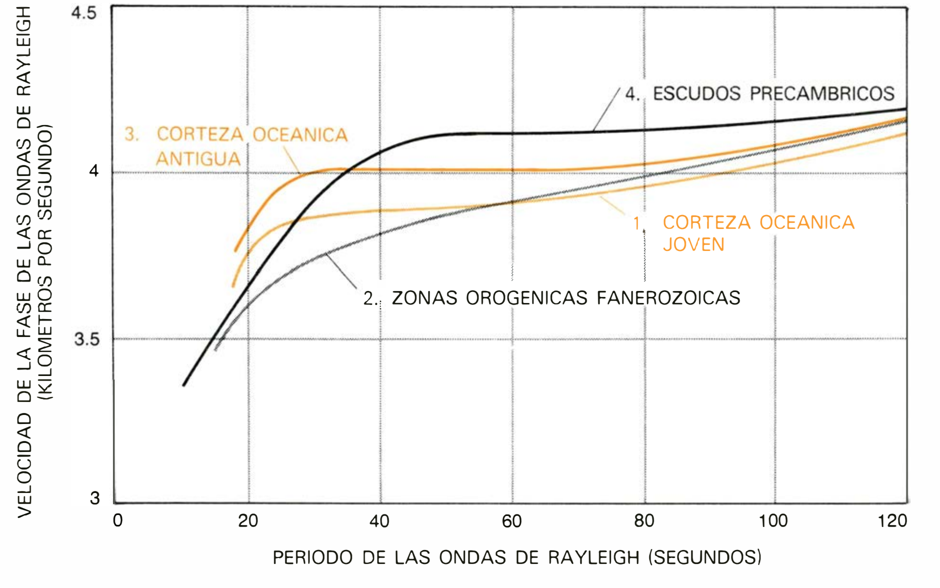
nia, cuyo análisis de la dispersión de ondas superficiales a través de bloques continentales estables le llevó a proponer que los contrastes estructurales entre los cratones y las cuencas oceánicas se extienden hasta una profundidad de 400 kilómetros como mínimo. Dziewonski había indicado, sin embargo, que los datos de la onda de superficie no permitían diferenciar, de forma concluyente, entre estructuras a estas profundidades; por lo que casi todos los geofísicos permanecieron escépticos sobre el particular.

Las pruebas que confirmaron, por fin, la idea de que las estructuras continentales son de un espesor sensiblemente superior a los 200 kilómetros no fueron aportadas por los experimentos de la onda superficial, sino por las observaciones realizadas con unas ondas mucho menos frecuentes, las *ScS*. Las ondas de cizalla se propagan casi verticalmente hacia abajo desde la fuente (un terremoto) y son reflejadas por la brusca discontinuidad que existe entre el manto sólido y el núcleo líquido.

La fase denominada ScS_1 (o simplemente *ScS*) es reflejada una vez por el contacto manto-núcleo hacia la estación receptora; la fase ScS_2 es reflejada dos veces por el núcleo y una vez por la superficie exterior, y así sucesivamente para las fases de orden superior. Las fases *ScS* tienen muchas propiedades que son de gran utilidad para el estudio de la heterogeneidad del manto superior. Por ejemplo, la diferencia entre los tiempos de propagación de ScS_2 y *ScS* es sensible a las variaciones de velocidad que la onda de cizalla tiene en el manto superior por debajo del punto en que ScS_2 se refleja en la superficie; pero no ocurre lo propio con las variaciones que tienen lugar cerca de la fuente y de la estación receptora, en donde ScS_2 y *ScS* siguen casi las mismas trayectorias. Por tanto, variando la localización del registro (y por ende la localización de los puntos de la superficie donde se reflejan las fases *ScS*), pueden detectarse heterogeneidades laterales en regiones donde las estaciones sísmicas están dispersas o faltan por completo, tal como ocurre en las cuencas oceánicas profundas.

En 1975 y 1976, primero en la Universidad de Princeton y luego en la Scripps Institution of Oceanography, Stuart A. Sipkin y yo mismo publicamos dos estudios de variaciones globales de velocidad en el manto superior basados en las fases *ScS*. Descubrimos que los tiempos de propagación medios de las ondas de cizalla con trayectorias verticales a través del manto y la corteza, por debajo las cuencas oceánicas, son unos dos segundos superiores a los tiempos de propagación correspondientes a los cratones. Puesto que se sabía, gracias a los datos de las ondas superficiales, que la zona oceánica de baja velocidad era más acentuada que su parte continental correspondiente, el mero hecho de que las ondas de cizalla se transmitieran de un modo más lento bajo los océanos no contituyó sorpresa alguna. No se esperaba encontrar valores elevados para la diferencia observada, ya que los modelos existentes sobre los contrastes estructurales entre los continentes y los océanos predecían diferencias mucho menores. Por ejemplo, los tiempos de propagación vertical de la onda de cizalla computados a partir de los modelos, realizados por Dziewonski, de la onda de superficie de las estructuras cratónicas y oceánicas diferían en menos de un segundo y no en cuatro segundos.

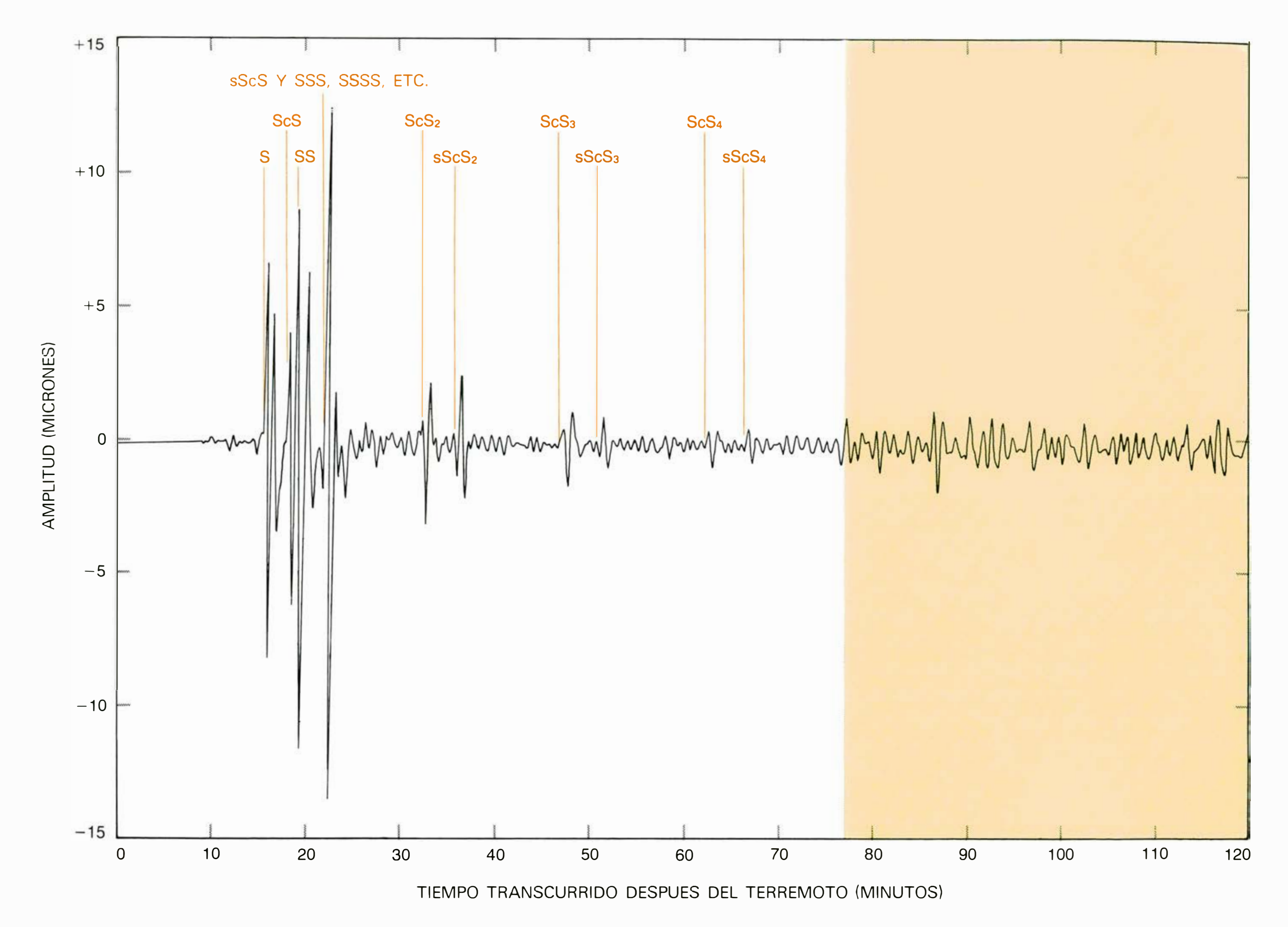
Estas grandes variaciones de los tiempos de propagación de *ScS*, combinados con los datos de la onda de superficie, implican la existencia de contrastes estructurales importantes entre océanos y cratones, contrastes que persisten hasta profundidades sin lugar a duda superiores a los 200 kilómetros, y que sobrepasan probablemente los 400. Las ondas superficiales son muy sensibles a las variaciones de los parámetros elásticos cerca de la superficie, mientras que las fases *ScS* permiten analizar el manto superior de manera casi uniforme a todas profundidades. Si se pretende construir modelos cratónicos y oceánicos donde los grandes contrastes, en punto a velocidad de la onda de cizalla requeridos por los datos de *ScS*, se concentren por encima de los 200 kilómetros, entonces, las curvas de dispersión de la onda de superficie, calculadas a partir de los modelos, entran en claro conflicto con las curvas de dispersión realmente observadas. Se puede demostrar con todo rigor que ningún modelo razonable, sujeto a esta restricción de profundidad, satisface adecuadamente los datos de *ScS* y los de la onda de superficie. Según parece, sólo los modelos con variaciones importantes por debajo de los 400 kilómetros, o a esa misma profundidad, pueden explicar de una manera satisfactoria ambos conjuntos de datos. Pueden construirse ese tipo de modelos [*véase la ilustración de la página 51*], pero no son únicos. En mi laboratorio y en otros se están realizando experimentos destinados a perfeccionar los modelos, y es de esperar que se obtengan pronto resultados más definitivos.



DEPENDENCIA DE LA FRECUENCIA de las velocidades de la onda de superficie. Tal dependencia puede emplearse para estimar la variación de la elasticidad del manto en función de la profundidad. Las curvas de dispersión características, aquí representadas, corresponden a regiones de la corteza con clasificación tectónica distinta. Para las ondas de superficie de baja frecuencia (es decir, las que tienen períodos superiores a unos 40 segundos) se observan velocidades máximas en las ondas que se propagan a través de los escudos continentales precámbricos (*curva 4*), mientras que las velocidades mínimas corresponden a ondas que se propagan por una corteza oceánica comparativamente reciente (*curva 1*).

La tectosfera continental

Aunque todavía no se puede puntualizar con mucha certeza cómo varían, en función de la profundidad, los contrastes entre la estructura elástica continental y la oceánica, y ni siquiera se puede concretar hasta qué profundidad máxima se extienden, sí está claro, sin embargo, que las divergencias (o contrastes) en la velocidad de la onda de cizalla se mantienen muy por debajo de los 200 kilómetros. Este solo hecho sólo obliga a reconsiderar una idea que ha sido de ca-



DISTINTAS ONDAS DE CIZALLA, cuyo movimiento se polariza normalmente a la trayectoria, representadas aquí por esta traza sismográfica compleja, cuyo registro se tomó en un observatorio situado en la isla de Oahu (islas Hawai) después del terremoto profundo localizado cerca de la isla de Tonga (Pacífico meridional) en octubre de 1974. Los diferentes tipos de ondas transversales responsables de las pulsaciones del sismograma, y designados por los símbolos en color de la parte superior, se explican en la ilustración de la página siguiente. En general, las ondas que llegaron al observatorio, después de 77 minutos de haber ocurrido el terremoto (*zona en color*), se propagaron siguiendo la larga trayectoria más larga alrededor de la tierra de la parte del círculo mayor que pasa por Tonga y Oahu. En ordenadas, la amplitud; en abscisas, el tiempo transcurrido tras el seísmo.

pital importancia en el modelo de la tectónica de placas: la idea de que las placas siempre están constituidas por litosfera.

El material litosférico es por definición resistente: puede aguantar las fuerzas de cizalla inducidas por las cargas que gravitan sobre la superficie, como por ejemplo montañas que actúan sobre ella durante millones de años, sin que se produzca una deformación permanente de cierta consideración. La determinación del espesor de la litosfera, a partir de observaciones sobre los fenómenos de carga y sobre el campo gravitatorio, fue emprendida por primera vez por los geofísicos a principios de siglo. Los modernos refinamientos instrumentales y los nuevos análisis que se aplican a los datos básicos han confirmado una de sus conclusiones originales: incluso bajo los cratones más estables, como los escudos fenoscándico y canadiense, el espesor eficaz de la litosfera no llega a los 200 kilómetros y, probablemente, no supera en mucho los 100. En realidad, el espesor medio de la litosfera cratónica parece no ser mucho mayor que el de una litosfera oceánica antigua. Así, por ejemplo, R. I. Walcott, del Ministerio Canadiense de Energía, Minería y Recursos, ha estimado que la litosfera del Canadá, probablemente representativa de las áreas continentales estables, mide unos 110 kilómetros de espesor, mientras que el espesor de la litosfera bajo las islas Hawai, y quizás en otras partes de las cuencas oceánicas antiguas, es superior a 75 kilómetros. El material existente bajo estas profundidades forma parte de la astenosfera; cuando en ésta actúan fuerzas de cizalla, se deforma rápidamente aunque éstas sean de poca monta.

En otras palabras, los datos sísmicos dan a entender que las regiones subcratónicas, con velocidades anormalmente elevadas de la onda de cizalla, se extienden por debajo la litosfera y dentro de la astenosfera. Las variaciones laterales profundas de los parámetros elásticos resultan ser notablemente coherentes con todas las atribuciones de edad de la corteza que pueden verse en la ilustración de las páginas 52 y 53. Caben dos explicaciones: o bien las estructuras continentales profundas, caracterizadas por velocidades altas de la onda de cizalla, están más o menos estáticamente acopladas a los continentes suprayacentes (es decir, están formando realmente parte de las placas continentales) o bien representan variaciones térmicas que, de algún modo, se mantienen dinámicamente gracias a flujos de masa en un manto en convección. Para que el anterior mecanismo fuera viable, todos los cratones tendrían que ser fundamentalmente lugares donde actuara un flujo descendente de material frío (ya que las bajas temperaturas son necesarias para explicar las elevadas velocidades de la onda de cizalla). Sin embargo, tales flujos tendrían unos efectos evidentes sobre el campo gravitatorio terrestre, que no

han sido observados hasta ahora. Además, una hipótesis de flujo de masa impone una serie de severas restricciones geométricas al campo de flujo que resultan incompatibles con otras restricciones impuestas a la dinámica del manto. Según parece, pues, las estructuras continentales profundas no pueden explicarse como elementos mantenidos dinámicamente por convección.

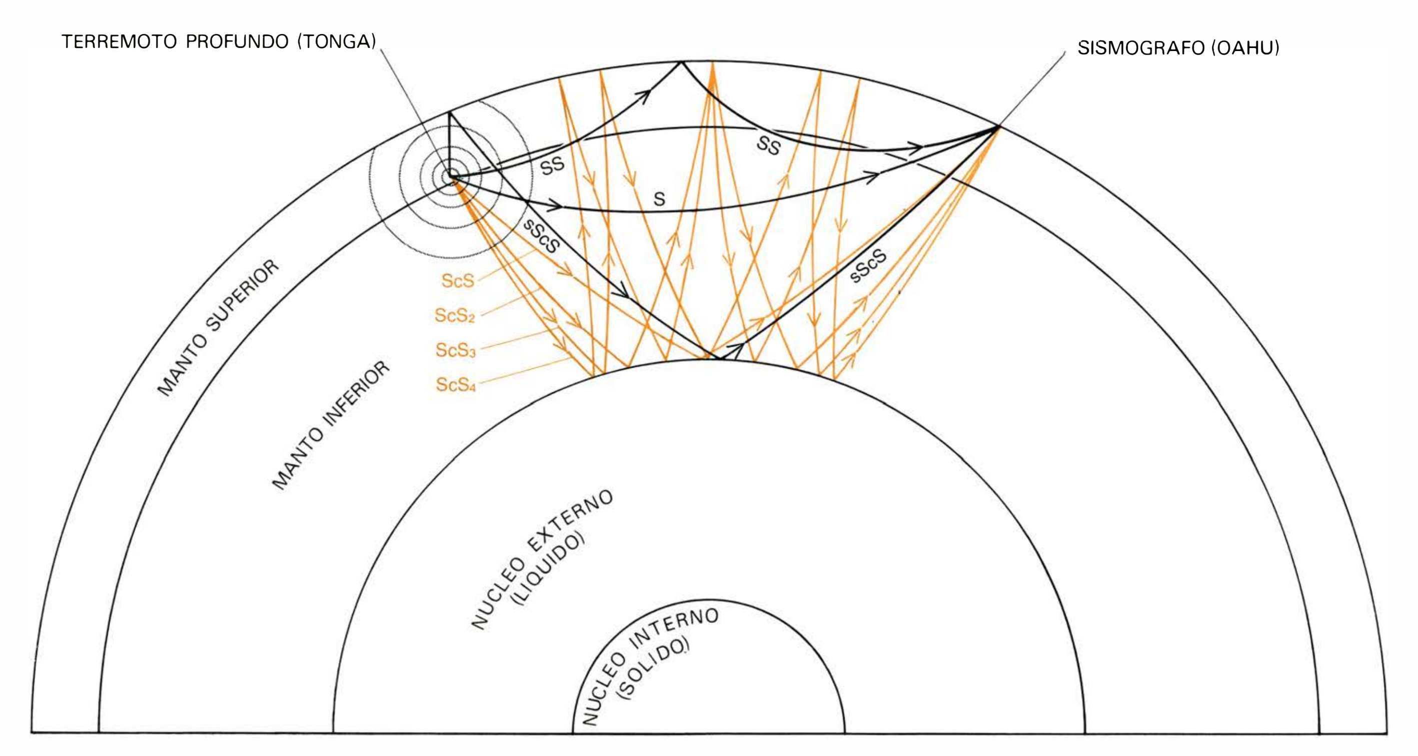
Por tanto, hay que concluir forzosamente aceptando que estas estructuras profundas constituyen las partes inferiores de las placas continentales que se han ido desplazando junto con la corteza durante centenares o incluso millares de millones de años, a pesar de que los materiales que forman dichas estructuras profundas sean parte de la astenosfera. Ante esta situación, algunos geofísicos han preferido simplemente volver a definir la litosfera, incluyendo en ella estas estructuras profundas. El uso inadecuado puede inducir a confusión; por ello he propuesto cambiar el término por el de tectosfera, que indica la región ocupada por las entidades coherentes que denominamos placas, manteniendo para la litosfera su definición clásica de capa de particular resistencia. (Este uso tiene precedentes en dos de los trabajos originales donde se esbozaba la teoría de la tectónica de placas. Walter M. Alsasser y W. Jason Morgan, de Princeton, aplicaron la palabra tectosfera en este sentido.)

La tectosfera se define según su comportamiento cinemático o, simplemente, por sus movimientos; por contra, la litosfera y la astenosfera se definen de acuerdo con su comportamiento dinámico, es decir, por el modo de responder a las fuerzas que se les imponen. A efectos prácticos, puede considerarse que, bajo los océanos, la tectosfera y la litosfera son idénticas en cuanto a extensión espacial, alcanzando espesores máximos de unos 100 kilómetros en las cuencas oceánicas más antiguas. Sin embargo, bajo los continentes, la tectosfera y la litosfera no son iguales, ya que la tectosfera cratónica se extiende por debajo de la litosfera, quizás hasta profundidades de 400 kilómetros o más.

Evolución térmica continental

Las variaciones de espesor en la tectosfera, deducidas a partir de datos suministrados por la onda de superficie y por la *ScS*, se correlacionan bastante bien con las estimaciones del flujo de calor procedentes de debajo de la corteza; ello sugiere que el espesor de la tectosfera está controlado por la estructura térmica del manto. El espesor de la tectosfera se correlaciona igualmente con la edad cortical; es decir, por regla general, la tectosfera más delgada está debajo de la corteza oceánica más reciente, mientras que la tectosfera más gruesa subyace a la corteza continental más antigua. En los océanos, el modelo de la tectónica de placas ha proporcionado una explicación simple a estas correlaciones. A medida que el nuevo fondo oceánico se expansiona lateralmente, alejándose del lugar de su formación, en la cresta de la dorsal oceánica, pierde calor por conducción hacia la superficie y, por tanto, las temperaturas que se hallan dentro de un "limite térmico estratiforme" superficial disminuyen con el tiempo. Así pues, el espesor de este límite estratiforme aumenta con la edad de la corteza, y disminuye el flujo de calor procedente de la superficie por él determinada. Los estudios realizados a través de modelos indican que el espesor de este límite térmico oceánico es, en esencia, igual al de la tectosfera y litosfera, con un valor medio de unos 70 a 100 kilómetros.

Los continentes muestran también una disminución sistemática de su flujo de calor superficial y un aumento del espesor tectosférico en función de la "edad cortical", siempre que la edad se defina como el tiempo contado a partir del último acontecimiento orogénico importante, no como el tiempo transcurrido desde que la corteza original se



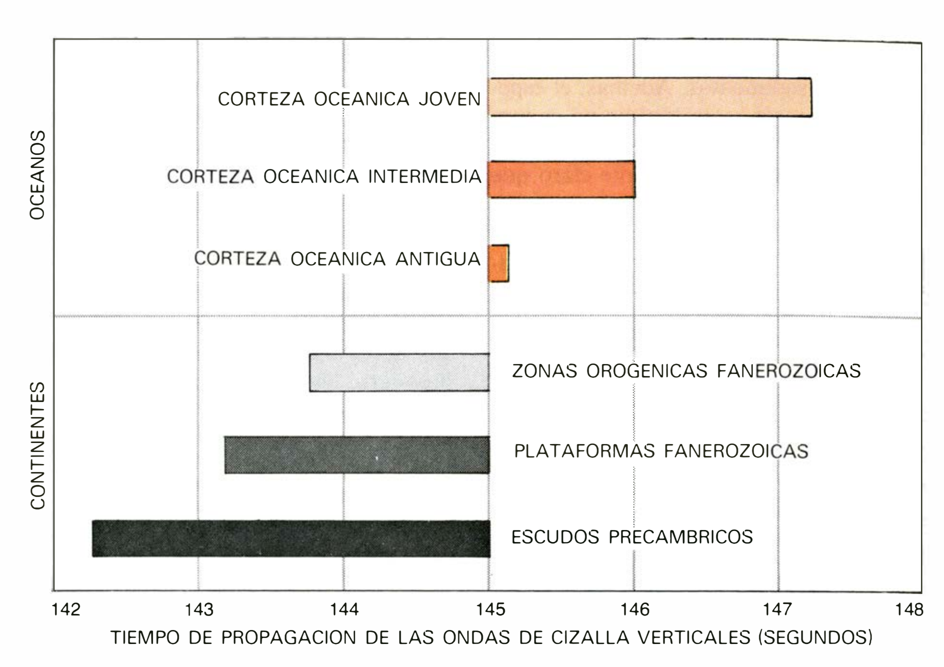
TRAYECTORIAS de las ondas de cizalla representadas en el sismograma de la página anterior. En este diagrama aparecen junto con sus denominaciones sismológicas convencionales. Las trayectorias *S* y *SS* indican ondas transversales que se propagan por el manto, yendo desde la fuente hasta la estación receptora sin alcanzar el núcleo terrestre. Las ondas que siguen la trayectoria *ScS* se reflejan una vez en la discontinuidad neta que separa el manto sólido del núcleo líquido antes de volver a la superficie. Las ondas que se propagan según las trayectorias denominadas ScS_2, ScS_3, etcétera, se reflejan más de una vez en la discontinuidad manto-núcleo, reflejándose igualmente en la superficie. Los múltiples tipos de ondas *ScS* son especialmente útiles para estudiar la estructura del manto superior a lo largo de las trayectorias recorridas entre la fuente y la estación receptora. A diferencia de las ondas de superficie, las de cizalla de esta clase son sensibles a las variaciones de elasticidad del manto a grandes profundidades. Proporcionan una información valiosa sobre la profundidad hasta donde se extienden ciertas diferencias estructurales entre la corteza continental y la oceánica.

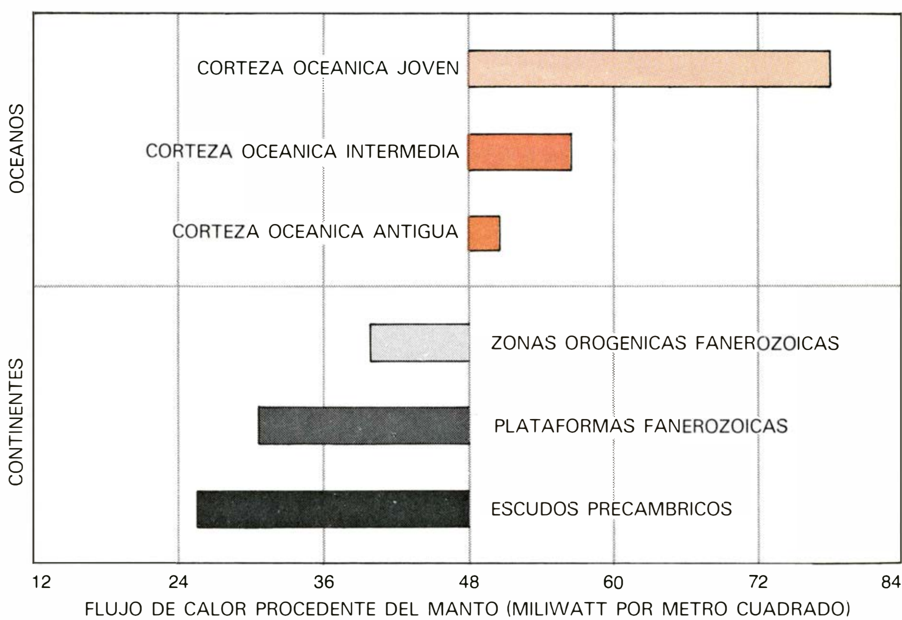
formó. Este comportamiento ha animado a algunos geofísicos a extender la hipótesis del límite térmico estratiforme a los continentes. Según esta hipótesis, el manto subcontinental se calienta intensamente durante las principales orogenias, enfriándose luego por conducción de calor hacia la superficie. Así pues, los cratones se interpretan como regiones que se han ido enfriando durante mucho tiempo y en donde el límite térmico estratiforme ha aumentado hasta alcanzar un espesor extremo. Los tiempos calculados para la duración de una degradación térmica de este tipo están dentro de lo razonable: el espesor de un límite térmico estratiforme que se ha ido enfriando durante uno o dos mil millones de años debería ser de casi 400 km; esto podría explicar los datos sísmicos.

A pesar de esta concordancia y el atractivo que pueda tener como teoría unificada de la evolución térmica de la tectosfera, existen serias dificultades cuando se aplica la hipótesis del límite térmico estratiforme a los continentes. El problema inmediato lo encontramos en las predicciones hechas por el principio de la isostasia. Puesto que la astenosfera no puede sufrir estáticamente importantes esfuerzos de cizalla, las presiones ejercidas sobre ella deben igualarse en el manto superior a una determinada profundidad mínima, denominada nivel de compensación. Para que el modelo resulte verosímil, la isostasia requiere que la masa contenida en columnas de idéntica superficie basal de océano, corteza y manto superior sean iguales, ya que la suma de los productos de los espesores de las capas y las densidades debe ser la misma.

Bajo los océanos, a medida que el manto se enfría y el límite térmico estratiforme crece, la densidad dentro de este límite aumenta. Para conservar la isostasia, la corteza y el manto superior experimentan una subsidencia, y aumenta la profundidad entre el nivel del mar y el fondo oceánico. De ahí que las rocas del manto se desplacen por debajo del nivel de compensación y sean sustituidas por agua, mucho menos densa, permitiendo que las columnas de masa permanezcan en equilibrio. Este mecanismo explica de forma muy exacta la morfología de las dorsales centro-oceánicas y la localización de los centros de expansión oceánica en las crestas de esas dorsales: el manto existente bajo la corteza acabada de formar es caliente y ligero, enfriándose y hundiéndose a medida que se aleja de los centros de expansión.

En los continentes, sin embargo, la historia de los movimientos verticales no concuerda tan bien con las predicciones de la hipótesis sobre el límite térmico estratiforme. Cuando la corteza continental y el manto se enfrían después de un episodio orogénico y térmico, tendrían que sufrir también una subsidencia que quedara registrada en los depósitos sedimentarios y fuera correlativa con ellos. (De otra forma, todos los cratones estarían bajo el agua). Aunque se observan grandes acumulaciones de sedimentos en algunos márgenes continentales y en determinadas cuencas del mismo tipo, los abundantes datos que se poseen sobre la historia de los cratones revelan algo muy distinto. Así, el crecimiento de un límite térmico estratiforme de 300 kilómetros de espesor (quizás el mínimo necesario para explicar los datos sísmicos de los cratones) predice una subsidencia cortical, y un consiguiente engrosamiento por sedimentación, de unos 20 kilómetros. Sin embargo, los cratones se caracterizan, de hecho, por su notable falta de cobertera sedimentaria, aflorando en los escudos enormes extensiones de rocas del zócalo, e incluso las plataformas están rara vez



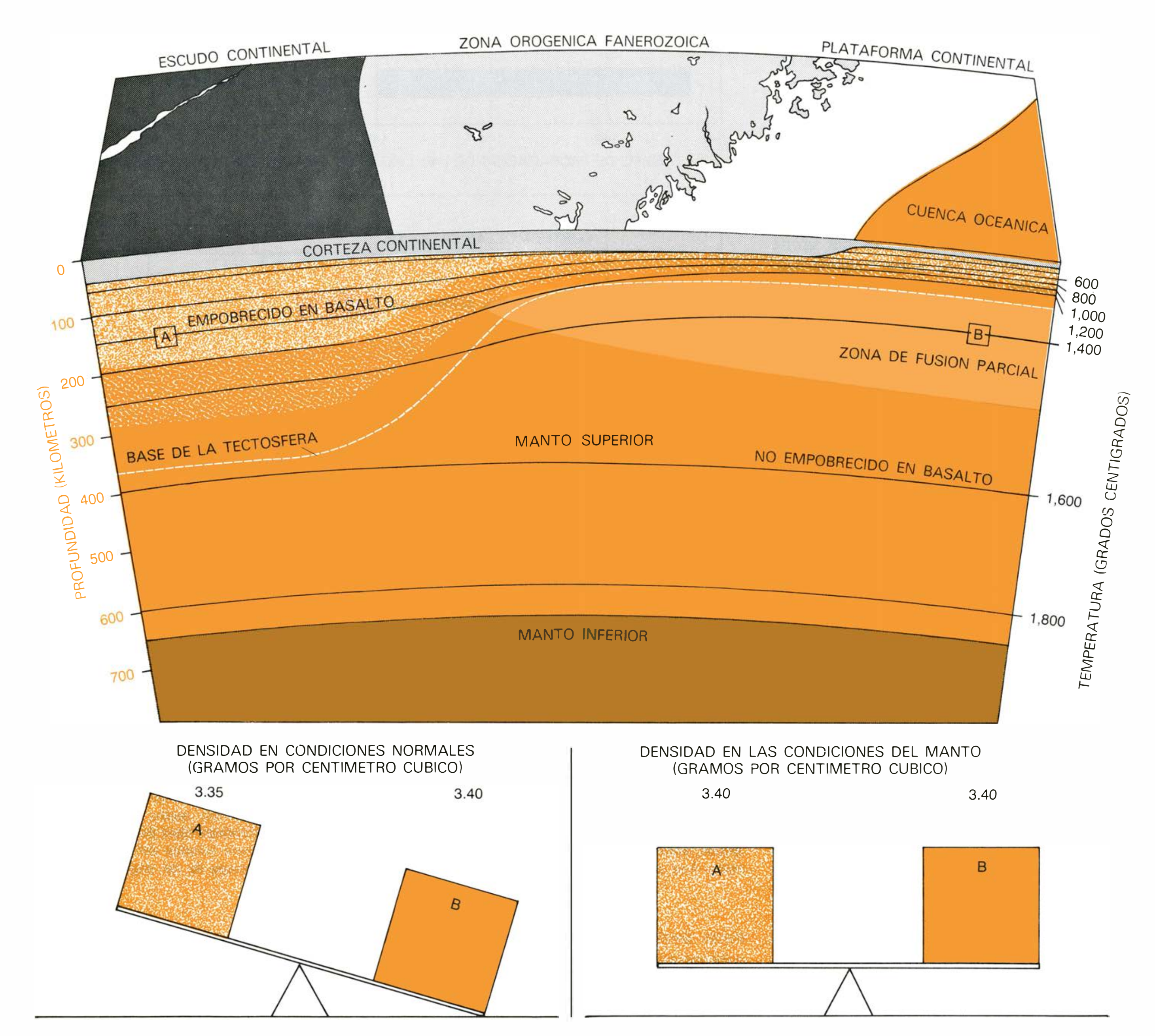


ESTRECHA CORRELACION observada, para diferentes tipos de rocas corticales, entre el tiempo que tarda una onda de cizalla en propagarse verticalmente desde una profundidad de 700 kilómetros hasta la superficie (*cuadro superior*) y la cantidad de calor que fluye del manto (*cuadro inferior*). Los tiempos de propagación verticales son promedios obtenidos a partir de gran número de tiempos de propagación de la onda múltiple *ScS*, siendo su valor medio en todo el globo de unos 145 segundos. Los valores característicos del flujo de calor que se dan son también promedios basados en el flujo de calor medido en distintos tipos de corteza, cuyo valor medio global es de 48 miliwatt por metro cuadrado. Esta suposición apoya la idea de que las diferencias sísmicas asociadas a diferentes tipos de corteza (continental, *abajo* y oceánica, *arriba*) provienen de diferencias en la estructura térmica del manto subyacente. Los colores se corresponden con los signos convencionales del mapa de las páginas 52 y 53.

cubiertas por más de unos pocos kilómetros de sedimentos. Además, el espesor de las cortezas cratónicas no supera el de las cortezas continentales más recientes. Queda, pues, bastante claro que la hipótesis del límite térmico estratiforme no sirve para explicar la existencia y desarrollo de la estructura continental profunda.

¿Qué explicación puede ofrecerse a cambio? En 1975 propuse que la evolución térmica de la tectosfera continental se hallaba íntimamente ligada a su evolución química, y que las estructuras continentales profundas, sugeridas por la sismología, constituían expresiones visibles de variaciones térmicas y de composición existentes en el manto superior. A esta hipótesis me llevó la necesidad de construir un modelo que contemplara una restricción especial: un modelo de una tectosfera subcontinental, estable a lo largo de grandes períodos de tiempo en un manto que debía suponerse inestable desde el punto de vista convectivo.

Voy a explicar, ahora, el por qué de este modelo. Una condición necesaria para la existencia de una estabilidad duradera bajo la litosfera es el equilibrio hidrostático, a saber: un estado de reposo donde las superficies de presión y densidad constantes coincidan y sean horizontales. La razón de ello estriba en que, en la astenosfera (por debajo los 100 kilómetros, más o menos), hasta unas pequeñas desviaciones del estado de reposo ideal harían fluir los materiales que la integran. Ahora bien, antes



DENSIDADES DE LAS ROCAS DEL MANTO, a determinado nivel bajo los continentes y océanos. Se consideran casi iguales en este modelo que propone el autor, si bien las temperaturas existentes bajo los primeros son inferiores. Para compensar un aumento de densidad debido a contracción térmica, postula que las rocas del manto tienen que hallarse diferencialmente empobrecidas en determinados componentes basálticos, los cuales están representados bajo las condiciones existentes en él por los minerales clinopiroxeno y granate. Por tanto, en condiciones normales las peridotitas que constituyen las zonas de raíces continentales frías serían menos densas que sus correspondientes partes suboceánicas. Por ejemplo, se estima que a una profundidad de 150 kilómetros, el manto subcontinental (*A*) posee una temperatura de unos 400 grados Celsius menor que el manto suboceánico (*B*), necesitándose un descenso de densidad, a través del mecanismo del empobrecimiento basáltico, de aproximadamente 0,05 gramos por centímetro cúbico. Las condiciones normales significan aquí 25 grados Celsius y presión de una atmósfera. Las condiciones del manto son temperaturas de 1000 y 1400 grados para las rocas del manto subcontinental y del manto oceánico, respectivamente, y una presión de 50.000 atmósferas para ambos. El diagrama de la parte superior es idéntico al de la página 51 con la substitución de las líneas de igual velocidad por las isotermas (*negro*).

indiqué que, para explicar las velocidades elevadas de la onda de cizalla características de las estructuras continentales profundas, se requieren temperaturas anormalmente bajas bajo los continentes y, por tanto, gradientes térmicos laterales en el manto existente bajo los continentes y océanos, muy por encima de los 100 kilómetros de profundidad. En un manto químicamente homogéneo, tales gradientes de temperatura inducirían gradientes laterales de densidad, y las superficies de densidad y presión constantes no coincidirían. En consecuencia, se produciría una inestabilidad convectiva, por lo que el flujo de material resultante desorganizaría con rapidez, y acabaría por destruir, las partes sublitosféricas de la tectosfera. Pero la tectosfera profunda continental no se halla desorganizada y destruida; muy al contrario, hay que destacar su prolongada estabilidad.

De ahí arrancó mi idea de que el manto no era químicamente homogéneo y que los gradientes térmicos laterales asociados con las estructuras continentales profundas se hallaban estabilizados frente a la desorganización convectiva producida por diferencias de composición. Se suponía que los gradientes de composición estaban dinámicamente ajustados, de suerte que la tectosfera subyacente a la litosfera estuviera en equilibrio hidrostático aproximado con el manto más caliente que la rodea. Es decir, que si el manto fuera químicamente homogéneo, existiría un exceso de masa bajo los continentes debido a las densidades más elevadas derivadas de temperaturas más bajas. De acuerdo con la hipótesis de una composición no homogénea, esa masa en exceso queda compensada localmente por deficiencias de masa controladas químicamente. Por decirlo de otro modo, las asociaciones mineralógicas existentes en el interior de las zonas de raíces cratónicas profundas tienen, en realidad, la misma densidad que las situadas al mismo nivel bajo los océanos, ya que las asociaciones subcratónicas son más frías. Serían menos densas si las temperaturas continentales y oceánicas fueran iguales.

Empobrecimiento basáltico

La deducción de que las estructuras continentales profundas difieren, en su composición, del manto que las rodea, se basaba sólo en razonamientos geofísicos, sin entrar en consideraciones geoquímicas específicas. Però las investigaciones posteriores revelaron pronto un mecanismo geoquímicamente atractivo y sencillo para generar heterogeneidades de composición: la densidad de la tectosfera continental podría rebajarse mediante la eliminación de un componente basáltico del manto.

El basalto es el magma más frecuente emitido en la superficie terrestre: cada año se eliminan del manto unos 20 kilómetros cúbicos de magma de este tipo, que pasan a engrosar la corteza. La mayor parte de este volcanismo se produce en los océanos a lo largo de los centros de expansión de las dorsales o en los "focos calientes", como las islas Hawai; además, las rocas de composición basáltica son también constituyentes importantes de la corteza continental. Cuando el basalto cristaliza cerca de la superficie, está fundamentalmente compuesto por los minerales plagioclasa (alúminosilicato sódico y cálcico) y clinopiroxeno (silicato de color obscuro, y composición química variable que contiene hierro, magnesio, calcio y sodio). Sin embargo, a profundidades subcorticales, el basalto cristaliza dando una asociación mineral denominada eclogita, que encierra también clinopiroxeno y posee, además, un mineral denso llamado granate, que representa la fase aluminosa en sustitución de la plagioclasa.

Hasta hace unos 20 años había geoquímicos que pensaban que el manto superior podía estar constituido en su totalidad por eclogita; ahora en cambio se cree que gran parte del mismo está formado por peridotita; ésta, a profundidades superiores a los 70 kilómetros, consta de una asociación de cuatro minerales: olivino (60 por ciento), ortopiroxeno (12 por ciento), clinopiroxeno (15 por ciento) y granate (13 por ciento). Esta clase de peridotita con cuatro fases recibe el nombre de lerzolita granatífera. Cada mineral de dicha lerzolita posee una temperatura de fusión diferente. Los componentes basálticos (clinopiroxeno y granate) funden a temperaturas más bajas que los componentes más refractarios (olivino y ortopiroxeno). Si la temperatura es suficientemente elevada, de suerte que parte del manto se funda (por ejemplo un 10 o un 20 por ciento), la fracción fundida tiene composición basáltica. Si luego se separa esta mezcla, lo que queda en la fracción residual es una roca empobrecida en basalto, constituida fundamentalmente por olivino y ortopiroxeno.

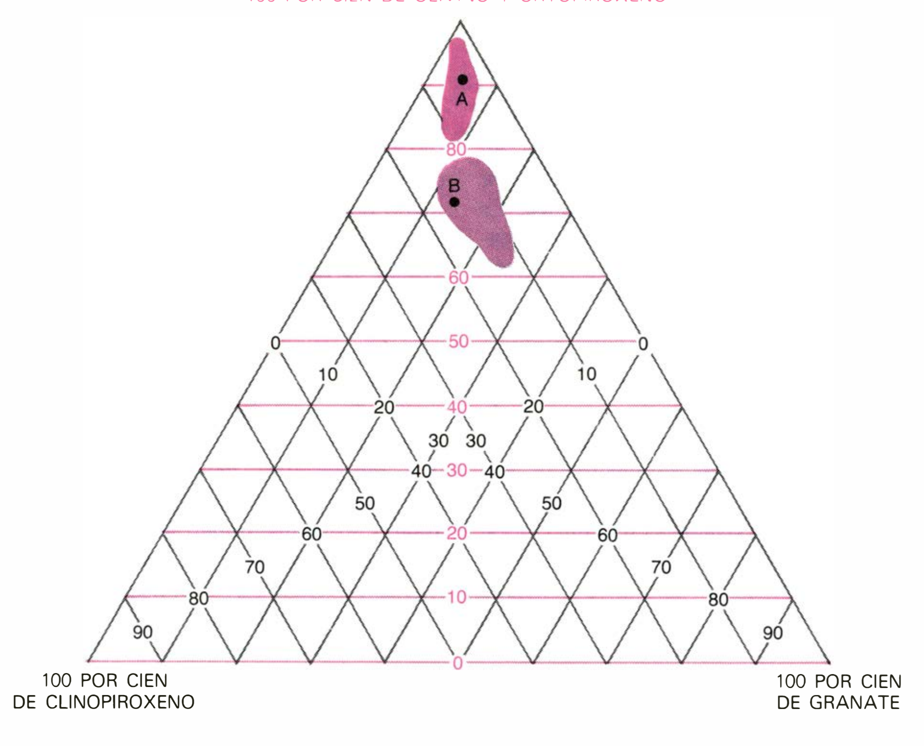
Lo realmente importante es que, a una temperatura y presión dadas, la densidad de esta roca residual empobrecida es algo inferior a la de su pariente próximo, la lerzolita granatífera; es decir: la separación de una componente basáltica del manto reduce ligeramente la densidad del material que lo compone según un valor característico de un 1 por ciento, más o menos. La razón por la que se produce esta disminución de densidad es fácil de comprender. La separación del basalto deja al manto con una proporción menor de granate, que es substancialmente más denso que los otros minerales de la lerzolita granatífera (3,7 gramos por centímetro cúbico, frente a 3,3 gramos por centímetro cúbico), disminuyendo también su contenido en hierro (el elemento abundante más pesado que en él se encuentra). Es sorprendente que estas relaciones de densidad no las tuvieran en cuenta los geofísicos y geoquímicos hasta hace muy poco. Los primeros resultados cuantitativos claros referentes a este problema se publicaron en un trabajo breve de Michael J. O'Hara, de la Universidad de Edimburgo, en 1975. La conclusión de O'Hara (el empobrecimiento en basalto hace disminuir la densidad del manto) ha sido confirmada más tarde por el trabajo experimental de Francis R. Boyd y R. H. McCallister, de la Carnegie Institution de Washington, y por mis copiosos cálculos numéricos.

Propuse en 1976 que las variaciones en composición requeridas para estabilizar la tectosfera continental venían inducidas por empobrecimiento basáltico de la tectosfera. Mis cálculos mostraron que el efecto debería ser cuantitativamente suficiente. A profundidades comprendidas entre 150 y 200 kilómetros, se estima que el manto subcontinental es más frío que el manto suboceánico, con una diferencia de temperatura de unos 300 a 500 grados Celsius. En un manto químicamente homogéneo, tales diferencias térmicas producirían diferencias de densidad que oscilarían entre 1 y 1,5 por ciento. Las diferencias de densidad de esta magnitud podrían compensarse con exactitud si la peridotita que ahora se encuentra a una profundidad de 150 a 200 kilómetros bajo los cratones, hubiera estado fundida en una proporción del 10 al 20 por ciento en algún momento de su historia, y si la fracción basáltica fundida se hubiera separado por migración hacia la corteza.

Esta hipótesis del empobrecimiento basáltico puede verificarse mediante observaciones geoquímicas directas. Rocas que se encontraban bajo los continentes, a una profundidad comprendida entre 150 y 200 kilómetros, se localizan hoy en la superficie, en forma de cantos y bloques redondeados, denominados xenolitos (rocas extrañas), y en el interior de formaciones conocidas con el nombre de chimeneas de kimberlita, que son los conductos erosionados de determinados volcanes especiales ubicados sólo en

regiones continentales estables [véase "Chimeneas de kimberlita", por Keith G. Cox; INVESTIGACION Y CIENCIA, junio, 1978]. Estas chimeneas han sido muy estudiadas, toda vez que se trata de la fuente esencial de todos los diamantes que se encuentran en la superficie terrestre (la localidad tipo de kimberlitas, roca volcánica de tipo peridotítico que rellena la chimenea, es la conocida zona productora de diamantes de Kimberley, en Africa del Sur). Los xenolitos, algunos de los cuales contienen diamantes, fueron arrancados, según parece, de las paredes del conducto volcánico durante el rápido ascenso del magma kimberlítico gaseoso que, procedente de gran profundidad, fue transportado hasta la superficie sin haber sufrido, la mayoría de veces, alteraciones químicas de importancia. Las investigaciones petrológicas y geoquímicas han demostrado que muchos xenolitos proceden de profundidades entre 100 y 250 kilómetros.

La mayor parte de los xenolitos que proceden del manto son lerzolitas con granates. Petrólogos como P. H. Nixon, de la Universidad de Papua (Nueva Guinea), y Boyd, de la Carnegie Institution, han apuntado que, al parecer, los xenolitos se hallan desprovistos de un componente basáltico; es decir: se caracterizan por contener menos granate y clinopiroxeno que el existente en la asociación mineral que se estima constituye la mayor parte del manto superior oceánico. Para demostrar este empobrecimiento, construí un modelo de la composición del manto superior cratónico, en el intervalo de profundidad 150-200 kilómetros, promediando los análisis químicos disponibles de los xenolitos de lerzolita granatífera originarios de las chimeneas de kimberlita. La naturaleza del empobrecimiento de esta lerzolita granatífera continental promediada, puede observarse en la ilustración inferior. A temperatura y presión normales, la densidad de este material continental resulta ser un 1,3 por ciento menor que la densidad que nos da el mejor modelo disponible del manto superior oceánico



PRUEBAS que sostienen la hipótesis del empobrecimiento basáltico. Se obtuvieron a partir de los xenolitos (literalmente "rocas extrañas") llegados hasta la superficie terrestre gracias a la erupción del magma basáltico de las chimeneas de kimberlita. Estas rocas son, seguramente, muestras representativas de la composición del manto superior existente bajo la corteza continental a profundidades variables entre los 100 y 250 kilómetros, y en ellas se han observado concentraciones de clinopiroxeno y granate (*mancha superior*) menores que las estimadas para el manto superior subyacente a la corteza oceánica (*mancha inferior*). La densidad de los xenolitos de kimberlita tiene un valor medio menor en unos 0,05 gramos por centímetro cúbico a la densidad estimada para la peridotita del manto superior oceánico, lo cual concuerda con el valor predicho por la hipótesis del empobrecimiento basáltico para la densidad media de la peridotita existente bajo la corteza continental. La composición de la peridotita empobrecida en basalto en el punto *A* de esta ilustración y en el correspondiente a la página 60 es: 67 por ciento de olivino, 23 por ciento de ortopiroxeno, 4 por ciento de clinopiroxeno y 6 por ciento de granate. La peridotita no empobrecida del punto *B* de ambas ilustraciones es: 60 por ciento de olivino, 12 por ciento de ortopiroxeno, 15 por ciento de clinopiroxeno y 13 por ciento de granate. Los componentes basálticos se indican en negro y los no basálticos en color. (Ilustraciones de Allen Beechel.)

(es decir, en la mitad del intervalo 1-1,5 por ciento requerido por la hipótesis del empobrecimiento en basalto).

El examen más detallado de las composiciones y densidades de los xenolitos de lerzolita granatífera proporciona una prueba más rigurosa de la hipótesis del empobrecimiento en basalto. Se supone que las diferencias de composición entre el manto continental y el oceánico son proporcionales a sus diferencias de temperatura; a su vez, se supone que las diferencias de temperatura disminuyen al aumentar la profundidad, desapareciendo cerca de la base de la tectosfera. El valor del empobrecimiento en basalto de las rocas continentales debería pues disminuir con la profundidad. Tal como anunciaran por vez primera Nixon y Boyd, esto es lo que ocurre de hecho. El contraste de densidad decrece con la profundidad, que es lo que prevé la hipótesis del empobrecimiento basáltico.

Esta hipótesis fue sugerida, indirectamente, por los datos que aportó la sismología: las elevadas velocidades de onda indican bajas temperaturas por debajo los cratones. El empobrecimiento en basalto puede desempeñar un papel más directo al explicar estas velocidades de las ondas sísmicas y, en especial, la velocidad de la onda de cizalla. Esto produce dos efectos importantes. En primer lugar, la separación de basalto disminuye el contenido en hierro del manto, y los experimentos han demostrado que la velocidad de las ondas sísmicas aumenta al decrecer el contenido en hierro. En segundo lugar, la separación de basalto eleva la temperatura de fusión del manto (ya que se eliminan los componentes que funden a baja temperatura), lo cual también aumenta la velocidad de las ondas, incluso a temperaturas por debajo del punto de fusión. Así pues, el empobrecimiento en basalto ayuda a explicar por qué las velocidades de la onda de cizalla resultan ser tan altas bajo los cratones.

En resumen, la hipótesis del empobrecimiento basáltico posee varios rasgos sugestivos: proporciona un mecanismo para estabilizar las partes sublitosféricas de la tectosfera continental frente a la desorganización convectiva, explica como mínimo dos caracteres relevantes de la petrología de los xenolitos kimberlíticos y es congruente con los datos sismológicos disponibles. Sin embargo, se necesitan muchas más pruebas antes de poder aceptar el modelo como base sobre la que construir una teoría del desarrollo continental. Las pruebas para verificar el modelo no se obtienen fácilmente, ya que los fenómenos en él representados están enterrados bajo cientos de kilómetros de roca. Y, sin embargo, los datos sobre la naturaleza de las zonas con raíces continentales están creciendo a un ritmo tal que si el modelo estuviera equivocado en sus grandes líneas nos enteraríamos muy poco de ello.

Nueva valoración de la estabilidad cratónica

Mientras tanto, quedan justificadas algunas especulaciones provisionales sobre la naturaleza de la estabilidad cratónica y continental. En el contexto del modelo por mí propuesto, la deriva continental es dinámicamente compleja; se trata de algo más que de una simple deriva pasiva de material cortical ligero sobre un trozo de litosfera, por otra parte indiferenciada. El modelo del empobrecimiento basáltico sugiere que el comportamiento tectónico continental se regula por variaciones químicas que se extienden hasta zonas profundas del manto.

Bajo las regiones de tectónica activa, en los bordes de placa o cerca de ellos (por ejemplo en el occidente de Estados Unidos o en Europa meridional), la capa de peridotita empobrecida parece ser delgada. Cuando el manto situado bajo esta capa delgada se enfría más allá de determinado punto crítico, se produce un hundimiento de material del manto no empobrecido provocando un retorno convectivo y el transporte de más calor hacia la superficie. La corteza y la parte más superior del manto de esta región son débiles y fácilmente deformados debido a su calentamiento por este retorno convectivo. Por contra, la capa empobrecida existente bajo los cratones se supone de un espesor mucho mayor, y la baja densidad de esta peridotita residual le impide mezclarse fácilmente de nuevo con el manto por convección, de tal forma que el material de la zona empobrecida queda estabilizado en un equilibrio casi hidrostático a temperaturas inferiores a las de su manto circundante. Puesto que su temperatura es menor y se ha eliminado su componente de temperatura de fusión baja, la viscosidad es más elevada y, por tanto, la zona de cizalla horizontal asociada con los movimientos de las placas queda confinada bajo la zona empobrecida. La corteza y la parte superior del manto quedan aislados de las perturbaciones térmicas y de los movimientos de masa asociados con alteraciones convectivas de pequeña escala, manteniéndose así la estabilidad cratónica.

Pero, ¿cómo se forma por vez primera una corteza subcortical gruesa, fría

y empobrecida? ¿De dónde proviene todo este basalto y a dónde va? Los geólogos han deducido que los procesos básicos de la tectónica de placas que se observan en la actualidad (expansión del fondo oceánico, deriva continental, subducción, etcétera) fueron activos durante la era precámbrica, que concluyó hace unos 600 millones de años. Algunos geólogos creen que existen pruebas de actividad de las placas tectónicas en rocas con una antigüedad de tres mil millones de años, o incluso mayor. Así pues, es razonable suponer que el desarrollo de la tectosfera ha sido gobernado por el ciclo de la tectónica de placas, como mínimo durante los últimos tres mil millones de años.

¿Qué procesos especiales pudieron ser importantes en el empobrecimiento? Hoy, el manto superior continental se está empobreciendo en basalto en dos tipos de zonas primarias: donde los episodios magmáticos están asociados con la apertura continental y en la cuña que forma el manto sobre las áreas de subducción. Las primeras se caracterizan por la formación de cantidades enormes de "basaltos de plataforma" extruidos durante los períodos de extensión cortical anteriores a la ruptura y la deriva de los continentes. (Por ejemplo, los basaltos de plataforma de la cuenca del río Paraná, en América del Sur, que fueron emitidos inmediatamente antes de la apertura del océano Atlántico meridional, tienen un volumen total de casi dos millones de kilómetros cúbicos. El volumen de manto empobrecido por este episodio debió ser como mínimo varias veces mayor.)

Más voluminosos aún son los basaltos emplazados en la corteza suprayacente a las zonas de subducción, desde las fosas oceánicas en dirección hacia el continente, a lo largo de los márgenes continentales "activos" y en los arcos de islas. Estas zonas están asociadas generalmente con el volcanismo andesítico más rico en sílice, si bien el basalto es, en realidad, el tipo de magma más frecuente, en particular en los arcos de islas recientes. Los petrólogos y geoquímicos están intentando identificar las fuentes de donde proceden estos magmas. Las andesitas pueden provenir de la corteza oceánica que ha sufrido subducción, pero la mayor parte del basalto se genera, por lo visto, a partir de la cuña de manto situada encima de la litosfera oceánica descendente. De ahí que una parte importante del manto residual empobrecido, e incorporado por fin a las gruesas zonas de raíces tectosféricas, pueda empobrecerse en un principio en regiones de formación rápida de arcos de islas, tal como ocurre hoy en el Pacífico occidental.

Pero la fusión y extracción de basalto son procesos confinados principalmente en el manto a una profundidad inferior a los 200 kilómetros. Por debajo de ella, la enorme presión reinante hace que rara vez se alcance la fusión requerida para producir basalto, si es que en realidad puede llegar a darse. Por tanto, una tectosfera empobrecida, de espesor superior a 200 kilómetros, no puede generarse sólo a través del proceso de fusión. E. R. Oxburgh y E. M. Parmentier, de la Universidad de Oxford, acaban de proponer otra fuente de origen. Han sugerido que parte de la peridotita empobrecida durante su ascenso y fusión en una dorsal centrooceánica, y que es transportada por expansión del fondo oceánico hacia una zona de subducción, puede, en el transcurso de esta última, unirse a la tectosfera.

Sea cual fuere la forma de generarse, toda tectosfera protocontinental existente en un ámbito de arco de islas, o a lo largo de un margen continental activo, sería de poco espesor, químicamente heterogénea y poco consolidada. El engrosamiento se hace con el tiempo. En primer lugar, las regiones dispersas del manto empobrecido y su corteza suprayacente se reúnen y sueldan por acreción a las masas continentales primarias, gracias al movimiento de las placas. Si bien estas últimas se hallan un tanto consolidadas, poseen aún una tectosfera delgada y, por tanto, pueden ser el asiento de una actividad orogénica. Buenos ejemplos de tectosfera continental que encuentran en esa etapa de desarrollo se localizan hoy en Indonesia, Canadá occidental y Patagonia.

Los episodios compresivos de mayor importancia que tienen lugar en los bordes de placa convergentes, y especialmente las etapas de colisión entre continentes, parecen producir una posterior consolidación y engrosamiento de la tectosfera. La historia de la corteza subsiguiente a una colisión continental ha sido analizada en detalle por John F. Dewey y Kevin C. A. Burke, de la Universidad estatal de Nueva York en Albany. Durante colisiones muy violentas, como la que se está produciendo entre la India y Asia, la corteza primitiva del margen continental se engrosa por compresión según un factor de dos, o superior, produciendo el levantamiento de un altiplano (como la meseta del Tibet donde los datos de la onda de superficie indican que la corteza tiene un espesor de unos 70 kilómetros). Dewey y Burke creen que este proceso de engrosamiento cortical compresivo implica una deformación a lo largo de una amplia zona. Esta deformación debería ser concomitante con un engrosamiento proporcional de la tectosfera subcortical producido por advección descendente de peridotita empobrecida. De esta manera podrían formarse las espesas zonas de raíces de los cratones.

Se cree, pues, que los episodios constructivos que inciden en el desarrollo de una tectosfera continental son: empobrecimiento, consolidación y engrosamiento. Para estabilizar una tectosfera potente en una región cualquiera, cabe que estos procesos hayan tenido que repetirse una y otra vez ya que, a menudo, deben enfrentarse a episodios destructivos producidos por inestabilidades fluidomecánicas del manto superior. En tanto que el empobrecimiento basáltico a gran escala es casi irreversible, tenemos certeza de que la consolidación y engrosamiento de la tectosfera no lo son. Todo calentamiento importante de una tectosfera fría, potente y estabilizada, producido por conducción de calor desde su periferia, gracias a fuentes térmicas internas o intrusiones, provoca este adelgazamiento y dispersión mediante el flujo lateral y ascendente de manto empobrecido de densidad baja. Estos movimientos pueden desempeñar un papel importante en las etapas de apertura y ruptura continentales. Un acontecimiento térmico de este tipo pudo desestabilizar las zonas antaño cratónicas del occidente de Estados Unidos, provocando la extensión y apertura generalizadas que hoy en día se producen en la Basin and Range Province, de los estados de Utah, Nevada y otros circundantes.

Los grandes rasgos de este esquema de la estructura y desarrollo de la tectosfera continental que aquí hemos presentado no son más que un modelo funcional, muy especulativo e incompleto, elaborado tras muchas generalizaciones discutibles y salpicado por los propios prejuicios del autor. Es necesario un examen más detallado para evaluar su viabilidad y comprobar la validez de las hipótesis en que se basa. Pero con independencia de su destino final, el modelo pone realmente de relieve la madurez a que han llegado hoy las ciencias de la tierra. Cualquier representación detallada de la estructura continental profunda y de su evolución debe plasmarse en la tela tejida con los innumerables hilos de la información geológica, geoquímica y geofísica. Sólo una representación así unificada nos revelará la verdadera naturaleza de los procesos dinámicos que han configurado los continentes.



Las trayectorias de los movimientos bruscos del ojo

Se expone cómo se consigue nueva información en torno al control que ejerce el cerebro humano sobre el movimiento ocular mediante un examen detallado de los giros bruscos y rápidos del globo ocular, que denominaremos sacudidas

A. Terry Bahill y Lawrence Stark

Cuando una persona lee, mira un cuadro, conduce un vehículo o incluso cuando observa fijamente un punto en la pared, sus ojos ejecutan una multidud de movimientos, la mayoría de los cuales son saltos tipo "staccato", que aquí denominaremos sacudidas. Los movimientos oculares de sacudida sirven para colocar en distintas zonas del campo visual un área diminuta situada en el centro de la retina que se llama fóvea. La fóvea es el lugar de la retina donde hay un empaquetamiento más denso de células receptoras visuales y, por tanto, es la parte más capacitada para permitir una visión de detalle. La información obtenida por las células de la periferia de la retina se utiliza para dirigir los diversos movimientos de sacudida, desplazamientos radiales cuya magnitud suele estar comprendida entre cuatro minutos y 15 grados. (Las desviaciones de la mirada de mayor ángulo se suelen llevar a cabo mediante una combinación de movimientos de la cabeza y de los ojos.)

Las trayectorias de los diferentes movimientos oculares de sacudida se hallan estrechamente relacionadas con los perfiles de actividad nerviosa de las motoneuronas que generan estos movimientos, perfiles que denominaremos señales de control. En el sistema físico donde se da el movimiento ocular, la carga que el globo en sí representa para los músculos es pequeña y constante; por otra parte, los músculos son fuertes y pueden contraerse con rapidez. De todo ello resulta que las señales de control neurológico que accionan los músculos se ven reflejadas con precisión en los movimientos del globo ocular. En otras palabras, la forma de las señales se puede deducir directamente de las trayectorias de las sacudidas. Cuando ya se llega a identificar las señales de control, se puede adquirir un conocimiento más amplio de la naturaleza de las sacudidas y del mecanismo que las genera.

En este artículo describiremos la aplicación de este método a varios tipos distintos de movimientos oculares de sacudida: sacudidas que se pasan (o no llegan a alcanzar) su blanco, sacudidas dobles y sacudidas de dirección oblicua. (A pesar de que la porción más extensa de la información que se ha recopilado acerca de los movimientos oculares se refiere a movimientos horizontales, la mayoría de los movimientos oculares tienen tanto una componente vertical como una horizontal.) Ahora bien, primero pasaremos a describir nuestro método de trabajo para la recolección de datos sobre los movimientos oculares y los instrumentos de los que nos hemos servido para el análisis de estos datos.

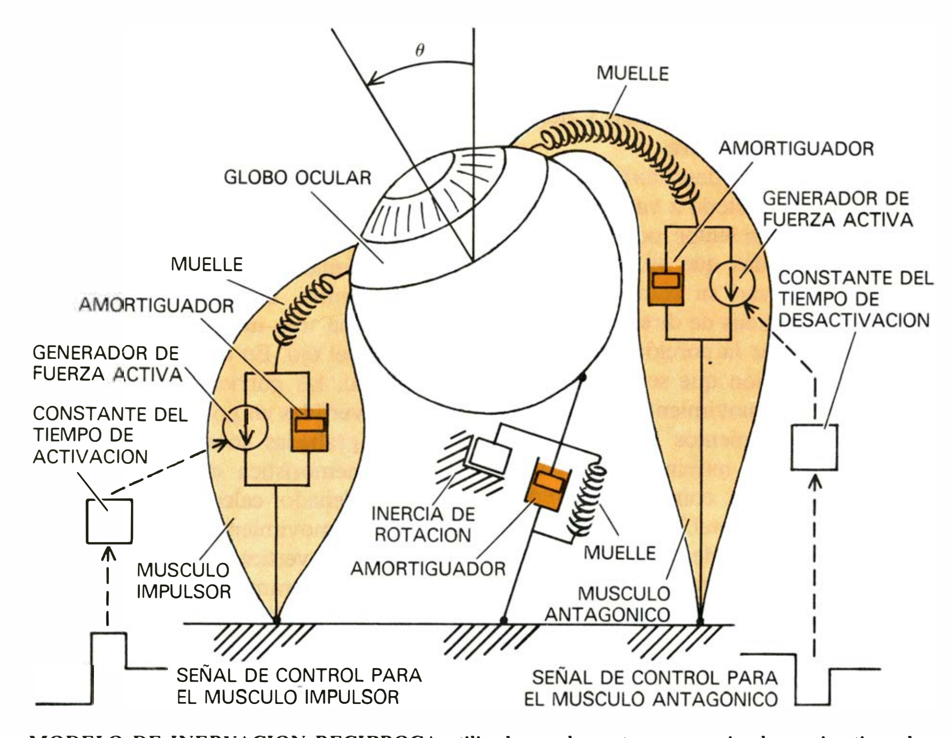
Para registrar los movimientos oculares, utilizamos un método fotoeléctrico. Se sitúa un par de fotodiodos sobre una montura de gafas de forma tal que estén orientados hacia el límite entre el iris, de color, y la esclerótica, blanca, a ambos lados de uno solo de los ojos del sujeto. Delante de éste, se coloca una pantalla sobre la cual se mueve el blanco (un puntito) de forma desordenada y brusca. Cuando, por ejemplo, en respuesta a un salto determinado del blanco, el ojo gira hacia la nariz, el fotodiodo que se halla más cerca de ella se ve expuesto a una mayor extensión de iris obscuro y a una menor área de la esclerótica clara, de modo que la corriente fotoeléctrica producida disminuye; simultáneamente, el fotodiodo que se halla más cerca de la sien se ve expuesto a más esclerótica y menos iris y su corriente se ve aumentada. La diferencia entre ambas corrientes nos proporciona una medición exacta de la posición del ojo. En nuestro sistema experimental, las corrientes fotoeléctricas eran convertidas en potenciales, amplificadas y grabadas sobre discos de una unidad memorística de ordenador. El propio ordenador calculaba las velocidades de los movimientos grabados. (Los movimientos verticales se midieron con un aparato ligeramente diferente.)

Suele ser corriente en los registros de movimientos oculares proceder a eliminar las frecuencias muy elevadas para evitar la interferencia de ruidos de fondo causados por los aparatos y otras irregularidades. No obstante, nosotros mantuvimos un margen de frecuencias muy amplio, de cero a 500 hertz (ciclos por segundo). Esta modificación es importante, pues el filtrar las frecuencias altas habría anulado los mínimos detalles de las trayectorias de las sacudidas, detalles de gran importancia, y habría dado lugar a velocidades calculadas de menor magnitud. Logramos reducir hasta un nivel muy bajo el ruido debido a los aparatos, de modo que siempre era mucho menor que la señal biológica que estábamos midiendo. Además, mediante un calibrado y ajuste continuos de los

LAS SACUDIDAS SIRVEN PARA MOVER la fóvea del ojo, es decir, el área pequeña situada en el centro de la retina y dotada de un gran poder resolutivo, y orientarla hacia diversos puntos del campo visual, como se describe en la secuencia de fotografías que aquí se presenta. Las trayectorias de estos giros del globo ocular, de tipo "staccato", se pueden visualizar iluminando el ojo de un sujeto con una luz brillante y concentrada. Mientras el sujeto fija su mirada en distintos puntos a lo largo de una línea horizontal (o incluso mientras mira fijamente un solo punto), el punto brillante de la córnea que refleja la luz constituye un punto de referencia útil para poner de manifiesto los movimientos oculares de sacudida generados. El comportamiento ocular típico de una persona con visión normal es ejecutar dos sacudidas por segundo. La mayoría de estos movimientos oculares tienen un orden de magnitud entre cuatro minutos y 15 grados de arco. Para desviaciones de mayor ángulo hay que mover la cabeza.

instrumentos, pudimos mantener una relación lineal entre la entrada y la salida de datos para sacudidas cuya rotación era de hasta 20 grados. El conjunto de nuestro dispositivo experimental no resulta de un gran costo y lo hemos utilizado recientemente, en colaboración con William F. Hoyt, de la Escuela de Medicina de San Francisco dependiente de la Universidad de California, y con B. Todd Troost, de la Escuela de Medicina de la Universidad de Pittsburgh, para la recopilación de información cuantitativa con el fin de mejorar el diagnóstico de los pacientes afectados por diversos trastornos del movimiento ocular.

El hecho de que nuestro montaje instrumental muestreara la trayectoria del ojo de modo más fino, más detallado, es decir, con mayor frecuencia, nos permitió una mayor aproximación a la identificación de las señales de control que accionan cada movimiento individual. Si se toman 10 muestras por segundo de la trayectoria (lo que corresponde a un margen de frecuencias de cinco hertz), se pueden estudiar los períodos de fijación entre las sacudidas, pero no así los movimientos rápidos que representan las propias sacudidas. Esta frecuencia de muestreo es válida para el estudio de los movimientos oculares en la lectura y para la descripción de los diversos caminos que sigue el ojo al pasear la mirada sobre diferentes panoramas u objetos. En cambio, con 100 muestras por segundo, podíamos obtener información acerca de la estrategia que utiliza el sistema visual para controlar la producción de movimientos oculares de sacudida. Por ejemplo, con este ritmo de muestreo, se observa que las sacudidas normales se dan a intervalos cortísimos, de unos 200 milisegundos. En 1962, Laurence R. Young, del Instituto de Tecnología de Massachusetts (MIT) y uno de los autores del presente trabajo (Stark) fueron los primeros en sugerir que el sistema de control para la generación de las sacudidas puede ser descrito como un sistema de control a partir de datos obtenidos por muestreo, es decir, un sistema de control en el que la salida de información se realiza de forma discontinua. Con una frecuencia de 1000 muestras por segundo, nuestros instrumentos nos proporcionaron una gran riqueza de informaciones cuantitativas detalladas acerca de los parámetros de las sacudidas, concretamente la velocidad máxima, la duración y la magnitud de la sacudida. Por otra parte, también permitieron observar las pequeñas variaciones en la forma de las sacudidas, algunas de las cuales se describen más adelante.



MODELO DE INERVACION RECIPROCA utilizado por los autores para simular varios tipos de movimientos oculares. Se basa en el concepto de que el giro del globo ocular se realiza mediante la inervación de un mínimo de dos músculos: uno impulsor, que se acorta y tira del globo ocular, y uno antagónico, que se alarga y relaja. En el modelo, las señales de control neurológico, o perfiles de emisión de impulsos por parte de las motoneuronas, como las que se muestran en la zona inferior izquierda y derecha, se consideran como entradas de información. Los elementos del modelo denominados constantes del tiempo de activación o desactivación reproducen el efecto suavizador que tiene el sistema ocular real sobre las señales mientras dichas señales llegan hasta, y son traducidas en movimientos por, los músculos impulsor y antagónico. La inercia del globo ocular para la rotación, como también la viscosidad y la elasticidad del sistema del globo y los músculos oculares, se representan mediante la inercia de la masa giratoria, el amortiguador y el muelle, situados por debajo del globo ocular. (Un amortiguador es un dispositivo viscoso que reduce la fuerza.) Los generadores de fuerza activa del modelo simulan la producción de fuerza muscular tal y como se origina en el interior del músculo real y no como se mide en el tendón del músculo. Cada generador de fuerza activa está conectado en serie con un muelle, el cual representa la elasticidad muscular. La salida de información del modelo es la posición del ojo θ, es decir, el ángulo que forma ésta después del giro con la posición de mirada al frente. Una propiedad del sistema ocular real es que las trayectorias del ojo constituyen reflejos fieles de las señales de control neurológico que originan los movimientos oculares. Utilizando el modelo para experimentar con señales de control de formas distintas, se lograron identificar las señales que generan determinadas variaciones concretas, observadas en las trayectorias de sacudida.

Con el fin de explicar las variaciones de las trayectorias de las sacudidas, se introdujeron dos nuevos dispositivos conceptuales. El primero utiliza una técnica estándar en ingeniería: cuando los bioingenieros desean estudiar un sistema relativamente complejo, suelen recurrir a un modelo para describir el sistema de un modo más sencillo. Los modelos son más simples y más manejables que los sistemas reales. Por ejemplo, un plano callejero es un modelo de las calles de una ciudad, y resulta más fácil dar instrucciones tomando al plano como referencia que atender al mundo real. Se ha desarrollado un gran número de modelos para facilitar el estudio del control neurológico del movimiento humano. Uno de los primeros modelos del sistema de control oculomotor fue ideado por René Descartes, el primero en darse cuenta de que los movimientos del globo ocular requieren la actividad coordinada de un mínimo de dos músculos: un músculo impulsor, que se acorta y tira del globo ocular proporcionando el impulso rotativo del movimiento ocular, y otro músculo antagónico, que se relaja y alarga. Por consiguiente, los movimientos oculares normales sólo se logran mediante la inervación recíproca de pares de músculos. Descartes sugirió este principio en 1626, pero hasta el año 1963 no fue incorporado a un modelo para movimientos oculares de sacudida, modelo que fue desarrollado por Gerald Cook, del MIT, y por uno de los autores del presente trabajo (Stark). Asociados con Michael R. Clark y Frederick K. Hsu, de la Universidad de California en Berkeley, pudimos desarrollar más tarde un modelo de inervación recíproca para todos los movimientos del ojo humano.

Los modelos de inervación recíproca son homeomórficos, siendo empleado este término en el sentido que le diera el biomatemático Richard E. Bellman: significa que existe una relación directa y detallada entre los componentes de cada uno de estos modelos y los componentes

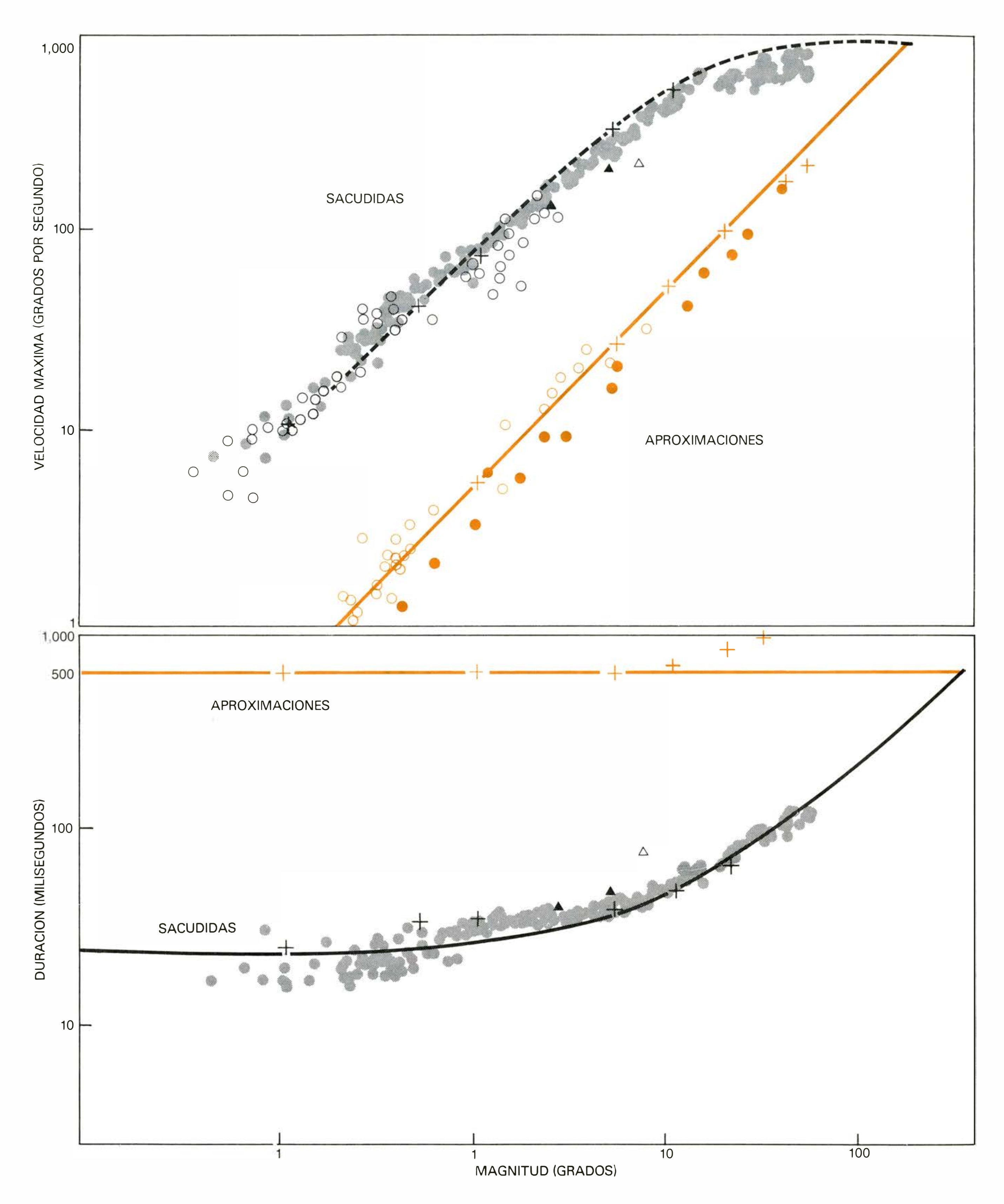
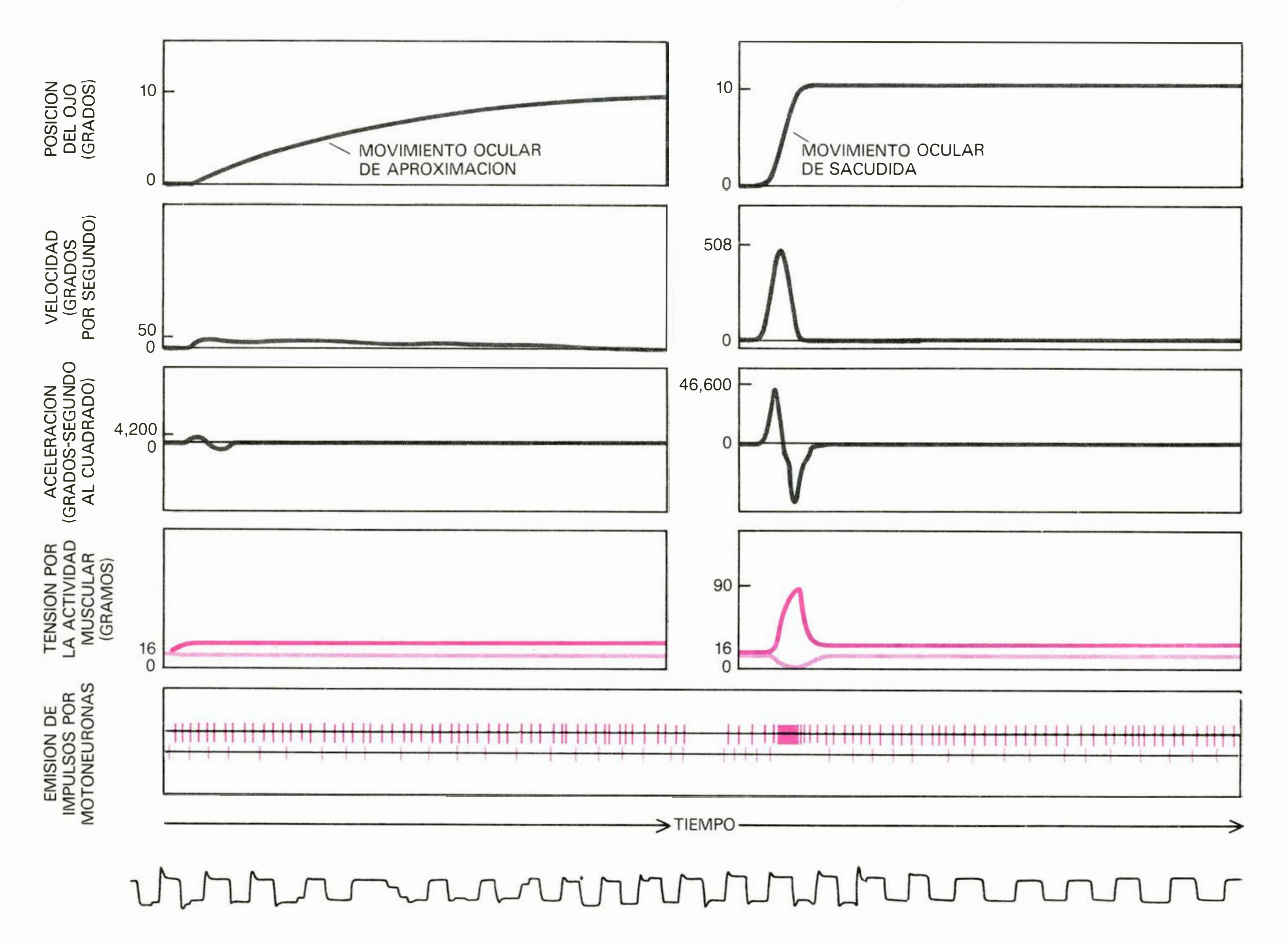


DIAGRAMA DE SECUENCIA PRINCIPAL para el ojo humano; representa la velocidad máxima (*arriba*) y la duración (*abajo*) de varios tipos de movimientos oculares frente a su magnitud. Los datos experimentales obtenidos en relación con los movimientos oculares se reagrupan en dos conjuntos principales en el diagrama: el de los movimientos rápidos, de tipo "staccato", de sacudida del ojo (*puntos grises*), que permiten la visión de detalle, y el de los movimientos oculares lentos de suave aproximación (*puntos de color*), que recorren la distancia entre los objetos alejados y los cercanos. Las sacudidas presentan variaciones; por ejemplo, algunas sobrepasan su posición final y vuelven a ella mediante un movimiento de sacudida rápido denominado sobrepasada dinámica (*círculos negros*). También puede volver el ojo a su posición final mediante un movimiento lento y deslizante denominado deslizamiento (*círculos de color*). En algunos casos, el ojo puede realizar dos sacudidas pequeñas poco espaciadas en lugar de una sacudida única y larga entre dos posiciones. Como nos indica el diagrama, el movimiento conjunto (*triángulos huecos*) no posee los parámetros típicos de una sacudida normal; más bien parece que son sacudidas los pequeños movimientos individuales (*triángulos compactos*). Las cruces negras del diagrama representan las sacudidas que fueron simuladas en el modelo de inervación recíproca; las cruces de color representan aproximaciones simuladas. Las líneas de trazos gruesos presentadas en la ilustración son funciones analíticas que se ajustan a la distribución de los datos experimentales.

del sistema real. La ventaja que comporta el trabajar con un modelo homeomórfico de los movimientos oculares de sacudida es que resulta posible no sólo simular una entrada de información en forma de señales neurológicas y una salida en forma de movimientos oculares, tal y como se puede hacer con un modelo de caja negra, sino también estudiar cómo queda afectada la salida por variaciones de los parámetros individuales del modelo y cómo dichos parámetros quedan afectados, a su vez, por las propiedades fisiológicas de los músculos y tendones reales del sistema de movimiento ocular.

Los parámetros de nuestro modelo de inervación recíproca representan la inercia del globo ocular, su viscoelasticidad al ser girado dentro de la órbita ocular, las elasticidades respectivas de los músculos y tendones oculares y la viscosidad aparente de los músculos oculares. (Una propiedad inherente al músculo es la relación entre la fuerza y la velocidad de la contracción, denominada relación de Fenn-Hill-Katz, y es tal que el mecanismo muscular que genera la fuerza parece estar tamponado como por un amortiguador o elemento viscoso que reduce la fuerza.) El modelo se construyó en un ordenador digital conectado a un osciloscopio de almacenaje (es decir, un osciloscopio dotado de memoria). Por tanto, en nuestras simulaciones de los movimientos del ojo humano podíamos variar la entrada de señales de control neurológico y observar inmediatamente el efecto de la variación sobre la salida de trayectorias de sacudidas. En concreto, pudimos experimentar el uso de señales de control de perfiles diferentes hasta encontrar la señal exacta generadora de una trayectoria con parámetros muy similares a los de un determinado tipo de sacudida observado.

Un aspecto importante en la construcción de un modelo es la validación del mismo. Nosotros llegamos a validar nuestro modelo por diversos caminos: de forma cualitativa, por comparación de los perfiles de las sacudidas reales con los del modelo de sacudidas; de forma cuantitativa, por comparación de los parámetros de las sacudidas reales con los del modelo; de forma analítica, por realización de un examen de la sensibilidad del modelo (variando sistemáticamente los parámetros individuales del modelo



SALTOS BRUSCOS DEL BLANCO. Tales saltos generan dos tipos de movimientos oculares: movimientos lentos de suave aproximación (*izquierda*) y movimientos agudos y rápidos de sacudida (*derecha*). Una aproximación se genera mediante una señal de cambio de ritmo que alcanza los músculos oculares: un aumento, moderado, pero mantenido, de la actividad nerviosa de las motoneuronas que inervan el músculo impulsor (*color obscuro*) y una disminución similar en la actividad nerviosa correspondiente al músculo antagónico (*color claro*). En este tipo de movimiento, lo que proporciona la fuerza impulsora que mueve lentamente al ojo hasta su nueva posición es la diferencia entre la posición inicial del ojo y la posición codificada por el nuevo nivel de equilibrio de la actividad nerviosa. Una sacudida se genera mediante una señal, compuesta por descarga y cambio de ritmo, que es transmitida a los músculos del ojo: un estallido de impulsos nerviosos, emitidos a alta frecuencia por las motoneuronas, que retrocede hasta un nivel de actividad nuevo, superior al inicial, en cuanto a las señales recibidas por el músculo impulsor (*color obscuro*), y una brusca caída en la actividad nerviosa seguida por un aumento hasta alcanzar un nivel de actividad, menor que el inicial, en cuanto a las señales recibidas por el músculo antagónico (*color claro*). En este caso la fuerza impulsora procede de la fuerte descarga y mueve al ojo con rapidez hasta su nueva posición; la señal de cambio de ritmo mantiene al ojo en su posición. Las irregularidades que se producen en las señales de tipo descarga y cambio de ritmo crean movimientos oculares de sacudida con trayectorias ligeramente variantes, lo cual se da en movimientos repetidos entre dos blancos separados 10 grados.

con el fin de observar su influencia en el conjunto del sistema del movimiento ocular); y de forma heurística, por simulación de movimientos oculares para los cuales no fue diseñado el modelo.

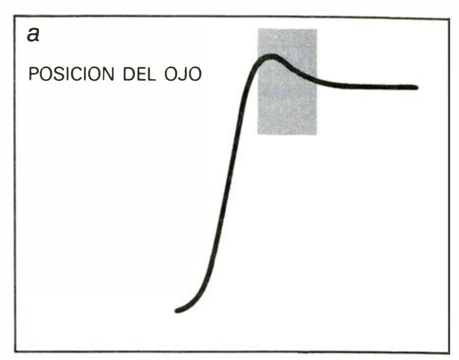
El concepto que hemos utilizado como segundo dispositivo está sacado del diagrama de Hertzsprung-Russell, empleado por los astrónomos para clasificar y comparar diversos tipos de estrellas. El diagrama es un gráfico que representa la luminosidad de las estrellas en función de su temperatura superficial o su color. La mayoría de las estrellas corrientes se encuentran situadas en el diagrama sobre una "secuencia principal": es decir, sobre la nube de puntos que cruza el diagrama en diagonal (véase "The New, or Modified, Form of the Theory of Stellar Evolution", de Henry Norris Russell: SCIENTIFIC AMERICAN, octubre, 1925). Hemos elaborado un diagrama similar que permite clasificar y comparar entre sí diversos fenómenos oculares de sacudida. En nuestro diagrama de secuencia principal, la velocidad angular máxima y la duración de los movimientos oculares se representan en función de la magnitud de los mismos [*véase la ilustración que aparece en la página 69*].

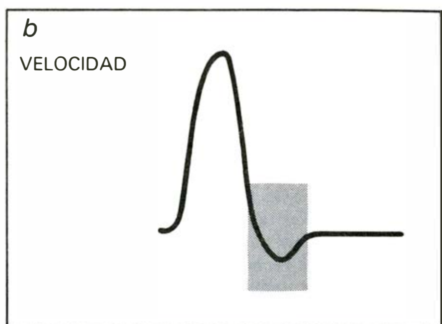
Los datos que hemos recogido de movimientos oculares se agrupan formando dos ramas en el diagrama de secuencia principal, una de ellas está constituida por los movimientos de sacudida rápidos y la otra por movimientos lentos de aproximación utilizados al recorrer con los ojos la distancia entre objetos lejanos y otros más cercanos. (Existen otros dos importantes tipos de movimientos oculares –los movimientos vestíbulo-oculares, que mantienen la estabilidad visual durante los movimientos de la cabeza, y los movimientos oculares de suave persecución, que siguen a un objeto en movimiento como puede ser un pájaro volando– pero éstos no aparecen en el diagrama de secuencia principal, ya que ninguno de ellos se manifiesta al exponer un sujeto a un blanco que da saltos.)

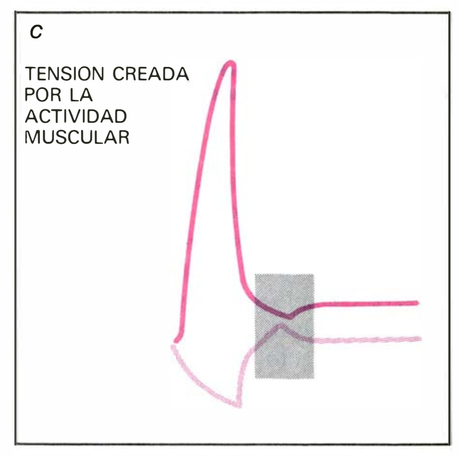
El diagrama de secuencia principal demuestra que es pequeña la variación que existe en los parámetros de los movimientos oculares de sacudida. Por ejemplo, en un sujeto normal descansado, el margen de duración de las sacudidas de 10 grados es de 38 a 45 milisegundos y el margen de la velocidad máxima de 420 a 520 grados por segundo. Del mismo modo, una sacudida que recorriera 10 minutos de arco tardaría entre 15 y 22 milisegundos y alcanzaría una velocidad máxima entre 11 y 15 grados por segundo. Estos márgenes son estrechos en comparación con los que se dan en otros sistemas biológicos. Nosotros mantuvimos márgenes de variación pequeños para los datos de secuencia principal actuando del modo siguiente: (1) desechando los datos procedentes de sujetos cansados; (2) representando los parámetros, no en función de la dimensión del movimiento del blanco, sino en función de la magnitud del movimiento ocular real, y (3) definiendo la magnitud del movimiento ocular de sacudida como la magnitud de la sacudida inicial, ignorando las derivaciones y ruidos posteriores con los que acaba el movimiento. Por consiguiente, aunque las sacudidas de sujetos humanos diferentes tengan parámetros ligeramente distintos, se puede decir que las sacudidas cuyos parámetros caigan fuera de los márgenes de secuencia principal suelen ser indicativas de alguna condición patológica.

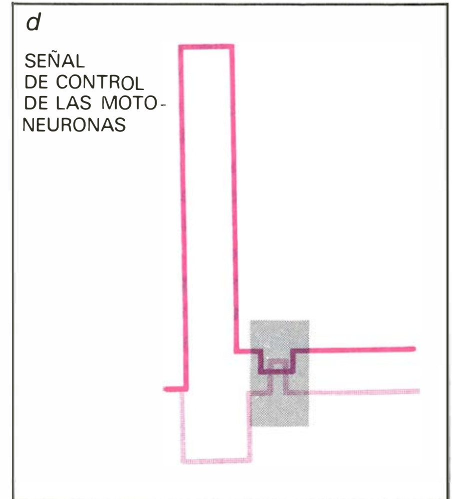
En 1965, el hallazgo de datos que, de un modo similar, seguían una secuencia principal llevó a Cook, Bert L. Zuber y uno de los autores del presente trabajo (Stark) a postular que no existían diferencias básicas entre los diversos movimientos de sacudida oculares. En otras palabras, los parámetros de todas las sacudidas se ajustan al diagrama de secuencia principal, independientemente de la magnitud de las sacudidas, de su finalidad y del estímulo que provocó su aparición. Los experimentos llevados a cabo ulteriormente han acabado de dar una base firme a esta teoría según la cual las sacudidas presentan una naturaleza unificada.

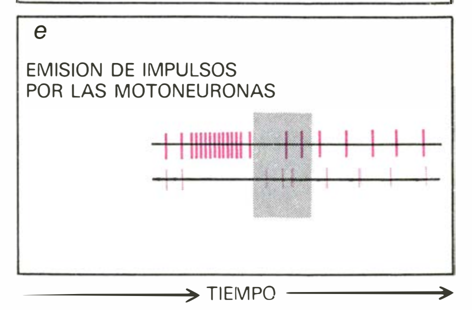
No obstante, los datos experimentales también nos revelan que, aunque los parámetros de secuencia principal y la forma general de las sacudidas varían



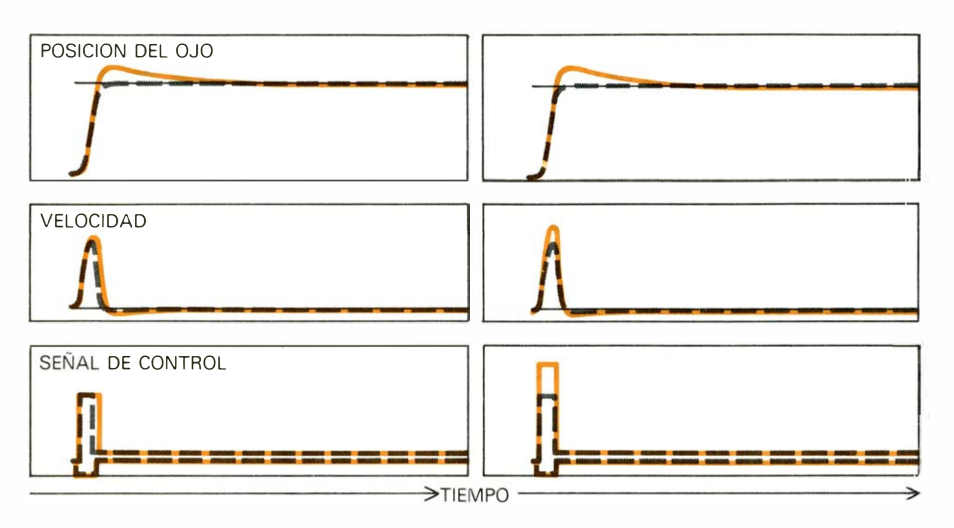








SIMULACION SOBRE EL MODELO de una sacudida de 10 grados. Pone en evidencia la variación de la trayectoria denominada sobrepasada dinámica: una sacudida normal que es seguida por un pequeño movimiento en la dirección opuesta, devolviendo al ojo a una posición que sobrepasó (*a*). Tal y como nos sugiere la relación entre la velocidad y el tiempo (*b*), la fase de retorno de una sacudida con sobrepasada dinámica es por sí sola una sacudida. La fuerza impulsora originaria de la segunda sacudida, de pequeña magnitud, es una inversión pasajera de los papeles del músculo impulsor (*color obscuro*) y del antagónico (*color claro*) hacia el final del movimiento ocular de sacudida (*c*). Este cambio en las tensiones creadas por las actividades de los músculos se origina a su vez por una descarga inversa en medio de la señal de descarga y cambio de ritmo que genera la sacudida (*d*), es decir, por una inversión pasajera de la actividad de las motoneuronas responsable del movimiento de sacudida. En la parte inferior se muestran los perfiles típicos de emisión de impulsos por parte de una motoneurona impulsora y de otra antagónica (*e*).



DESLIZAMIENTOS LENTOS que se dan al final de una sacudida; los tales pueden ser resultado de una magnitud excesiva del componente de descarga (*color*) de una señal de control normal en cuanto a otros aspectos (*líneas de trazos negra*). Cuando la señal de descarga es demasiado intensa (*derecha*), bien porque se ha reclutado un número excesivo de motoneuronas para la emisión de impulsos, bien porque emiten los impulsos a una frecuencia excesivamente alta, la diferencia entre la posición del ojo al final de la descarga y la posición del ojo codificada por la señal de cambio de ritmo genera un movimiento lento y suave que sigue a la sacudida primera y devuelve al ojo a la posición que sobrepasó. El mismo tipo de movimiento se origina cuando la emisión de impulsos por parte de las motoneuronas dura un tiempo excesivo y la descarga resulta demasiado larga (*izquierda*). Las líneas delgadas negras marcan la posición final del ojo y la velocidad. El análisis cuantitativo de las simulaciones sugiere que en los sujetos con visión normal abundan con mayor frecuencia los errores de amplitud de la descarga.

poco, la estructura de las sacudidas varía mucho, es decir, que existe un gran número de pequeñas desviaciones respecto a la trayectoria normal de la sacudida. Por ejemplo, algunas sacudidas se pasan de su objetivo, otras no completan el movimiento con una única sacudida estándar, etcétera. Estas desviaciones aparecen con relativa irregularidad, variando de un día a otro y de un movimiento ocular a otro; incluso a menudo son monoculares, es decir, el efecto se limita a un solo ojo. (De hecho, las variaciones que aparecen en las trayectorias de sacudidas de los dos ojos de un sujeto suelen ser bastante diferentes.) La explicación de este fenómeno aparentemente paradójico –es decir, de la presencia de variaciones de la trayectoria en movimientos oculares de parámetros estereotipados que siguen una secuencia principal– se halla examinando con más detalle los movimientos oculares de sacudida.

Hasta el momento, hemos considerado las sacudidas como unidades, pero en realidad cada sacudida está compuesta por varios componentes dinámicos o movimientos oculares menores. Lo que produce la variación en los movimientos oculares de sacudida es la variación en el número y en la ordenación de estos componentes menores. Hemos sido capaces de determinar que todas las variaciones en las trayectorias de sacudida –todos los cambios en las configuraciones de los componentes dinámicos de las sacudidas– se deben a variaciones en las señales de control neurológico de dichas sacudidas. (En las personas que tienen los músculos del ojo debilitados, es evidente que las variaciones tienen otro origen.)

Resultan ser una cuestión mal entendida los motivos por los que se producen señales de control algo irregulares respecto a las normales; su aparición no depende de la amplitud de las sacudidas, de la naturaleza o visibilidad de los blancos, de las condiciones generales de iluminación ni de la finalidad de las sacudidas. Lo que sí sabemos es que la fatiga es responsable de algunas desviaciones, y parece ser que la tensión psicológica también podría tener un papel. En cualquier caso, las señales de control varían y originan variaciones que se observan en las trayectorias de las sacudidas. Como ya hemos explicado antes, las trayectorias reflejan las señales con tal fidelidad que puede observarse con precisión cómo los cambios en las señales afectan la forma de las sacudidas.

En primer lugar, consideremos las señales de control que generan movimientos oculares de forma normal. Los movimientos de sacudida y los de aproximación suave, que constituyen las dos ramas del diagrama de secuencia principal, se generan por dos tipos distintos de señales de control neurológico. Un movimiento lento y suave de aproximación (sobre todo un movimiento divergente en que ambos globos oculares giran en sentidos opuestos) es el resultado de un cambio de ritmo en la frecuencia conjunta de emisión de impulsos por las motoneuronas que inervan los músculos oculares: consiste en un aumento, mantenido a baja frecuencia, de la actividad de las motoneuronas del músculo impulsor y de una disminución similar para el músculo antagónico, de modo que se establece un nuevo estado de equilibrio en cuanto al nivel de emisión de impulsos. La fuerza impulsora del movimiento ocular es la diferencia entre la posición inicial del ojo y la posición ocular que viene codificada por el nuevo nivel de actividad. En cambio, un movimiento de sacudida se genera mediante una señal que consta de dos partes, una fuerte descarga de impulsos y luego un cambio de ritmo respecto al nivel de actividad anterior. En este caso, la fuerza impulsora del movimiento ocular es la fuerte descarga de impulsos, un breve estallido de impulsos emitidos a una frecuencia muy elevada por las motoneuronas del músculo impulsor y una pausa paralela en la actividad de las motoneuronas del antagónico. La descarga sirve para mover el ojo con rapidez de un punto a otro; la señal de cambio de ritmo que le sigue sirve para mantener al ojo en su nueva posición. Una señal de descarga mueve al ojo con una velocidad muy superior que la atribuida a una señal de cambio de ritmo: por ejemplo, un movimiento de suave aproximación de 10 grados dura unos 500 milisegundos, es decir, unas 10 veces más que una sacudida de la misma magnitud.

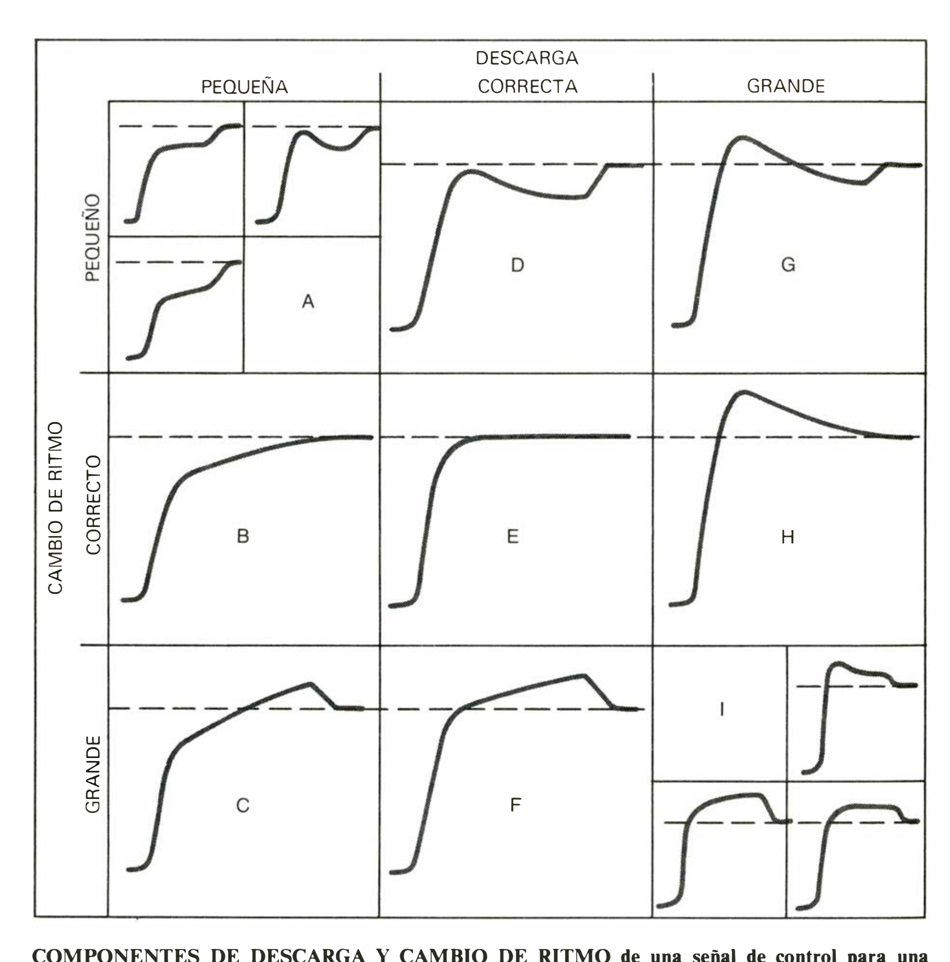
Para llegar a entender el modo en que las irregularidades de las señales de control compuestas por descarga y cambio de ritmo crean las diminutas variaciones en las trayectorias de las sacudidas, consideremos el siguiente ejemplo: la variación denominada sacudida con sobrepasado del blanco, en la que el ojo viaja más allá de lo que será su punto de detención final, volviendo luego a él y deteniéndose. La mayoría de las sacudidas del ojo humano manifiestan este comportamiento de un modo u otro. Existen tres maneras diferentes en que el ojo puede variar su trayectoria de sacudida sobrepasando el blanco; se distinguen por el modo en que el ojo vuelve a girar en sentido opuesto hasta la posición final. En la sobrepasada dinámica, se da una fase de retorno rápido; por ejemplo, en una sacudida en que el ojo se ha pasado en un grado, el retorno suele emplear 20 milisegundos, alcanzando el globo ocular una velocidad máxima de aproximadamente 60 grados

por segundo. En la sobrepasada deslizante, se da un retorno lento y deslizante que dura más de 200 milisegundos, alcanzando una velocidad máxima de cinco grados por segundo si el ojo se ha sobrepasado en un grado. En la sobrepasada estática, el ojo permanece fijado entre 150 y 200 milisegundos en la posición última que no coincide con el blanco, hasta que finalmente un efecto de retroinformación visual provoca una sacudida correctora. La sobrepasada estática es el único tipo de esta variación que puede percibirse a simple vista y, sin duda, por ello es el observado con mayor frecuencia por los expertos en clínica. Un exceso de este tipo de variación suele ser el síntoma de un trastorno del cerebelo.

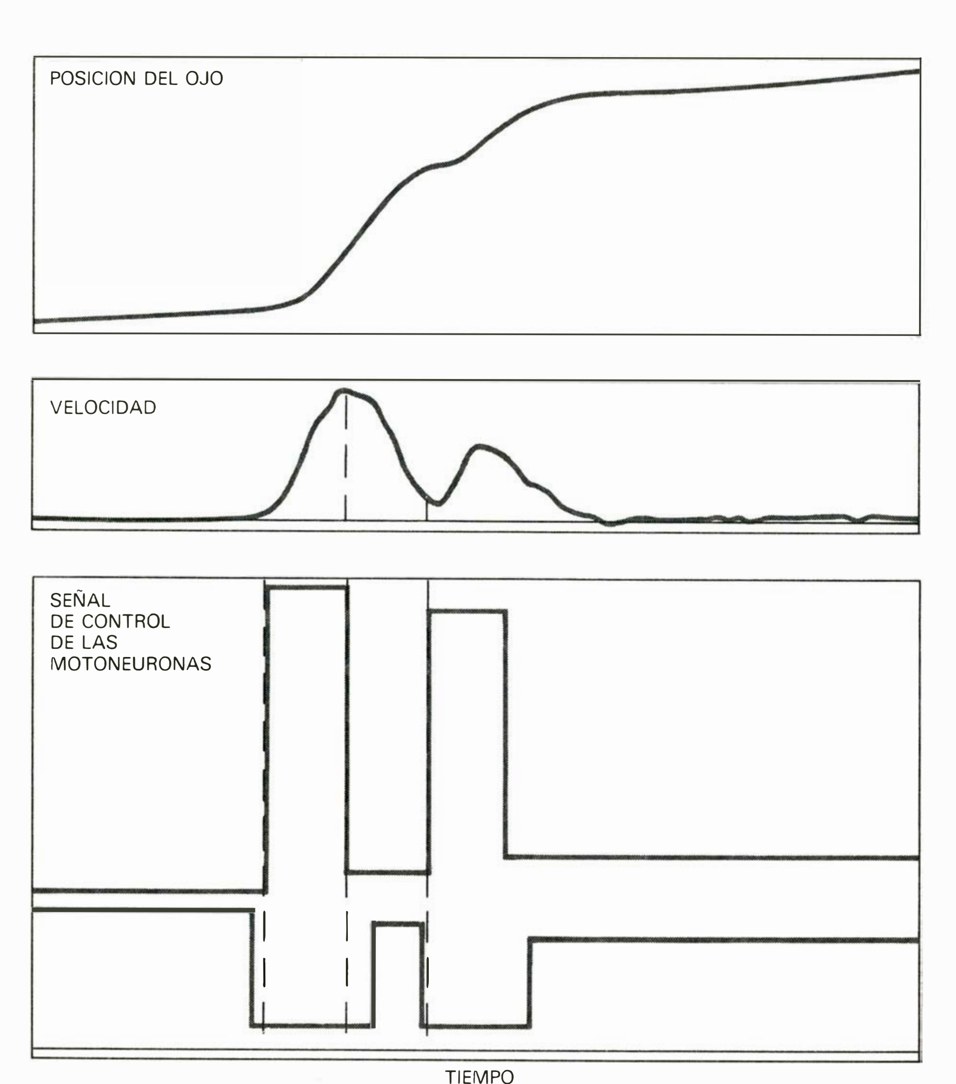
La mayoría de las sacudidas presentan sobrepasadas del tipo dinámico. Al parecer, esta irregularidad se manifiesta de forma caprichosa, de tal modo que un sujeto puede evidenciar esta variación en la gran mayoría de sus sacudidas en un día determinado y presentar ésta tan sólo en escasas sacudidas al día siguiente, mientras que en sujetos diferentes la forma que toma la variación puede variar muchísimo. Parece que la sobrepasada dinámica es independiente de la magnitud de la sacudida; hemos registrado sacudidas de todas dimensiones (desde pocos minutos de arco hasta 50 grados), unas acompañadas de sobrepasada dinámica y otras carentes de ella. Es más, la sobrepasada dinámica no parece limitarse a unas condiciones oculares iniciales ni a una dirección de movimiento; esta variación se ha detectado en sacudidas que se inician con la mirada al frente, pero también en otras en las que el ojo se halla girado 35 grados de esta posición al iniciarse el movimiento; se ha observado en sacudidas en dirección a la nariz y en otras que se alejan de ésta. La magnitud de la sobrepasada es mayor cuanto mayor es la magnitud de la sacudida, pero en una sacudida corriente, de unos 10 grados, la sobrepasada dinámica suele darse con un retorno típico de 0,25 grados. Si comparamos las fases de retorno de las sacudidas que presentan sobrepasada dinámica con los datos de la secuencia principal, vemos que poseen los mismos parámetros de secuencia principal que las fases de retorno de las sacudidas cortas.

El sistema ocular –los músculos y el globo oculares y el conjunto del tejido circundante– está muy amortiguado, como si en todas partes del sistema existieran fuertes amortiguadores que enlentecieran el movimiento y eliminaran las oscilaciones del globo ocular. Por tanto, las fases de retorno rápidas que se dan en las sacudidas con sobrepasada dinámica, y que podrían parecer oscilaciones, no pueden hallar su origen en la mecánica interna del sistema; se originan a nivel neurológico. Ya que las fases de retorno tienen un perfil similar al de las sacudidas, podría ser que fueran generadas por señales de sacudida de las motoneuronas. Con el fin de aclarar esta posibilidad, hicimos simulaciones de sacudidas con sobrepasada dinámica.

Las simulaciones ponen de manifiesto que una sacudida que presenta sobrepasada dinámica se genera cuando existe una inversión de la actividad de las motoneuronas –es decir, otra descarga de impulsos suplementaria pero inversa– en el transcurso de una señal de control estándar para sacudidas del tipo descarga y cambio de ritmo. Para el músculo impulsor, la señal de control comprende un aumento brusco de la actividad de las motoneuronas seguido por una súbita disminución y posteriormente por un nuevo aumento hasta alcanzar el nuevo nivel de equilibrio determinado por la señal de cambio de ritmo. Por el contrario, la señal para el músculo antagónico consiste en una disminución aguda de la emisión de impulsos (el fenómeno que originariamente causa la relajación del músculo) seguida por un incremento brusco y luego por una nueva caída hasta el nuevo nivel de equilibrio. Como consecuencia de estos perfiles de actividad neurológica, la dirección de la fuerza generada por las tensiones creadas por la actividad de ambos músculos se ve invertida en medio del movimiento de sacudida, haciendo que el globo ocular sea



COMPONENTES DE DESCARGA Y CAMBIO DE RITMO de una señal de control para una sacudida; dichas componentes pueden estar desajustadas de muchas maneras, dando lugar a 13 trayectorias posibles, que se muestran aquí. Por ejemplo, cuando el cambio de ritmo es correcto pero la descarga es demasiado grande (*H*), se genera una sobrepasada deslizante. Cuando la descarga es demasiado pequeña (*B*), la sacudida resultante se detiene a poca distancia de su posición final y el movimiento es completado mediante un deslizamiento. Cuando el componente de cambio de ritmo es incorrecto (bien demasiado fuerte o demasiado débil), la sacudida mueve al ojo hasta una posición que no corresponde al blanco, posición en la que permanece el ojo hasta que una retroinformación visual induce una sacudida correctora, bien hacia adelante (*D*), bien hacia atrás (*F*), pero alcanzando, por fin, el punto del blanco. Cuando tanto la descarga como el cambio de ritmo son excesivamente grandes (*I*) o pequeños (*A*), la sacudida primaria puede estar seguida por un deslizamiento hacia la derecha, hacia la izquierda o por ningún deslizamiento, dependiendo de las dimensiones relativas de los componentes. La línea de trazos de cada trayectoria indica la posición final del ojo, o sea, la que corresponde al blanco.



SACUDIDAS QUE SE SOLAPAN (*arriba*). Se trata de dos sacudidas normales espaciadas un intervalo tan corto de tiempo que sus perfiles de velocidad se solapan (*centro*). Las simulaciones realizadas sobre el modelo indican que este solapamiento no se da en las señales de las motoneuronas que generan las sacudidas (*abajo*). En la mayoría de las sacudidas, la descarga concluye a medio camino del movimiento ocular, de tal modo que en las sacudidas solapadas existe una caída de la actividad de las motoneuronas entre ambas señales. Los sujetos cansados presentan a menudo sacudidas que se solapan.

nuevamente arrastrado hacia una posición que sobrepasó, deteniéndose en este punto. (La tensión creada por la actividad de un músculo es la fuerza desarrollada por el músculo, no tal y como es medida a nivel del tendón muscular, sino tal y como se origina en las interacciones proteicas en el interior del músculo.)

¿Pero qué tiene de particular la señal de control que invierte la actividad de las motoneuronas creando sacudidas con sobrepasada dinámina? Clark y uno de los autores del presente trabajo (Stark) han demostrado –utilizaremos la terminología de los ingenieros referente al control de las condiciones óptimas– que la señal de control para una sacudida, de tipo descarga más cambio de ritmo, es temporalmente óptima, es decir, que no existe forma alguna para una señal de control que sea capaz de mover al ojo con mayor rapidez entre dos puntos del campo visual que la forma exacta utilizada por el cerebro para las sacudidas. Datos electromiográficos obtenidos recientemente en sujetos humanos por Alan Scott, Robert Kenyon y uno de los autores del presente trabajo (Stark) han verificado que, efectivamente, las inversiones de las señales de control predichas por el modelo de inervación recíproca se observan en los seres humanos. La generación de una sacudida con sobrepasada dinámica es una cuestión que ha de ser decidida por el cerebro con relativa antelación, es decir, antes de que se constituya la señal. No obstante, hemos comprobado que no todas las variaciones de la trayectoria de las sacudidas se forman con esta antelación; por ejemplo, en la sobrepasada deslizante, se genera cuando existe un error en la constitución de la señal de control compuesta por descarga y cambio de ritmo.

Recordemos que un movimiento ocular de sacudida con sobrepasada deslizante se compone de una sacudida normal que trasciende su posición final y de un movimiento de lento retorno deslizante hasta volver a ésta. Este movimiento de retorno lento fue denominado deslizamiento por Robert Daroff y sus colaboradores, de la Universidad de Miami. Las sobrepasadas deslizantes son mucho menos frecuentes que las dinámicas en las sacudidas que efectúan los ojos de sujetos normales descansados. No obstante, esta frecuencia aumenta considerablemente con la fatiga y en algunos estados patológicos del ojo. En los casos en que se puede desechar la fatiga como factor de control, el análisis de los deslizamientos oculares puede ayudar a diagnosticar la oftalmoplejía internuclear, síndrome que se observa a menudo en la esclerosis múltiple.

Nuestro estudio de los movimientos oculares deslizantes se inició analizando el ajuste de estos deslizamientos al diagrama de secuencia principal. Pudimos observar que la velocidad y la duración de los deslizamientos dependen de su magnitud y que ambos parámetros mencionados vienen a situarse en la rama de movimientos de suave aproximación del diagrama. Por tanto, parecía probable que, al igual que dichos movimientos de aproximación suave, fueran el resultado de la fuerza generada por una diferencia entre la posición inicial del ojo y la posición codificada por un cambio de ritmo en la emisión de impulsos de las motoneuronas. Realizamos simulaciones de sacudidas con sobrepasada deslizante en el modelo de inervación recíproca, con el fin de averiguar esta posibilidad.

Las simulaciones llevaron a pensar que un deslizamiento es el resultado de un desajuste de los componentes de fuerte descarga y de cambio de ritmo de la señal de control que inicia la sacudida. Por ejemplo, se da una sobrepasada deslizante cuando el componente de cambio de ritmo de la señal de control es correcto con respecto a la distancia que separa el punto de mira de un blanco determinado y, por el contrario, el componente de la descarga es demasiado intenso (bien porque es excesivo el número de motoneuronas reclutado para la emisión de impulsos, bien porque las frecuencias con que emiten los impulsos son exageradamente elevadas) o demasiado amplio (al ser excesivamente larga la duración de la descarga). Ocurre entonces que la descarga impulsa el ojo hacia una nueva posición, situada más allá del blanco; pero como

la señal de cambio de ritmo es la correspondiente a la posición del blanco, el ojo efectúa una ligera deriva volviendo a la posición de equilibrio. Del mismo modo, si la descarga de la señal es demasiado pequeña, el ojo se detiene a una corta distancia de su posición correspondiente al blanco, y luego se ve lentamente arrastrado hacia adelante por la señal de cambio de ritmo. Este último fenómeno se puede denominar "infrapasada" deslizante.

Conviene señalar que los movimientos oculares deslizantes no son movimientos de suave aproximación. A pesar de sus semejanzas, ambos tipos de movimientos difieren en varios aspectos. Por ejemplo, los movimientos de suave aproximación constituyen un fenómeno binocular y, en cambio, los movimientos deslizantes suelen ser monoculares. Además, existe cierto lapso de tiempo entre un salto del blanco y el movimiento de aproximación correspondiente; dicha pausa no se da entre el momento que se produce el error en la sacudida y el movimiento deslizante que lo corrige.

Son numerosas las formas que pueden revestir los desajustes de ambos componentes de la señal de control. En una sacudida dada, puede ser que cualquiera de ambos sea inferior, superior o igual a su magnitud correcta. Si lo incorrecto es el componente de cambio de ritmo, por ser demasiado elevado o demasiado bajo para una posición del blanco, se establece un error de posición relativo al nivel de equilibrio; el ojo permanece en la posición que no corresponde al blanco durante centenares de milisegundos hasta que una retroinformación visual provoca una sacudida correctora. Cuando ambos componentes son excesivos, por demasiado elevados o por demasiado bajos, el movimiento ocular de sacudida resultante presenta un deslizamiento hacia la izquierda o la derecha, o puede no presentar ningún movimiento deslizante, según cuál sea el componente con mayor error.

Los deslizamientos tan sólo son un tipo de los diversos movimientos variantes de la trayectoria de las sacudidas que ven aumentada su frecuencia por la fatiga. La fatiga crea irregularidades en las señales de control de las sacudidas, de tal modo que éstas pueden enlentecerse y llegar a romperse en movimientos múltiples. Cuando, por ejemplo, un sujeto ha realizado unos 500 movimientos de sacudida de 10 grados, el punto de fijación del ojo se alcanza con una imprecisión mayor, la cual exige un mayor número de sacudidas de corrección. Las sacudidas aberrantes, es decir, aquellas cuyos parámetros no se ajustan a la secuencia principal del diagrama, empiezan a hacer su aparición. En algunos casos, el ojo alcanza la posición final no con una sacudida amplia sino con dos sacudidas menores denominadas sacudidas dobles. Al llegar a este nivel de fatiga, si se le pide al sujeto que realice sacudidas "más precisas", puede conseguirlo. Tras efectuar unas 1200 sacudidas de 10 grados, el sujeto aún es capaz de realizar algunas sacudidas de forma normal, pero sus velocidades máximas son inferiores en un 10 por ciento a las correspondientes a las primeras sacudidas del sujeto aquel mismo día. Entonces es cuando empiezan a aparecer las sacudidas solapadas: son sacudidas dobles tan próximas entre sí que sus perfiles de velocidad están solapados. Todas las sacudidas que el sujeto pueda efectuar posteriormente a este período manifiestan claros indicios de fatiga.

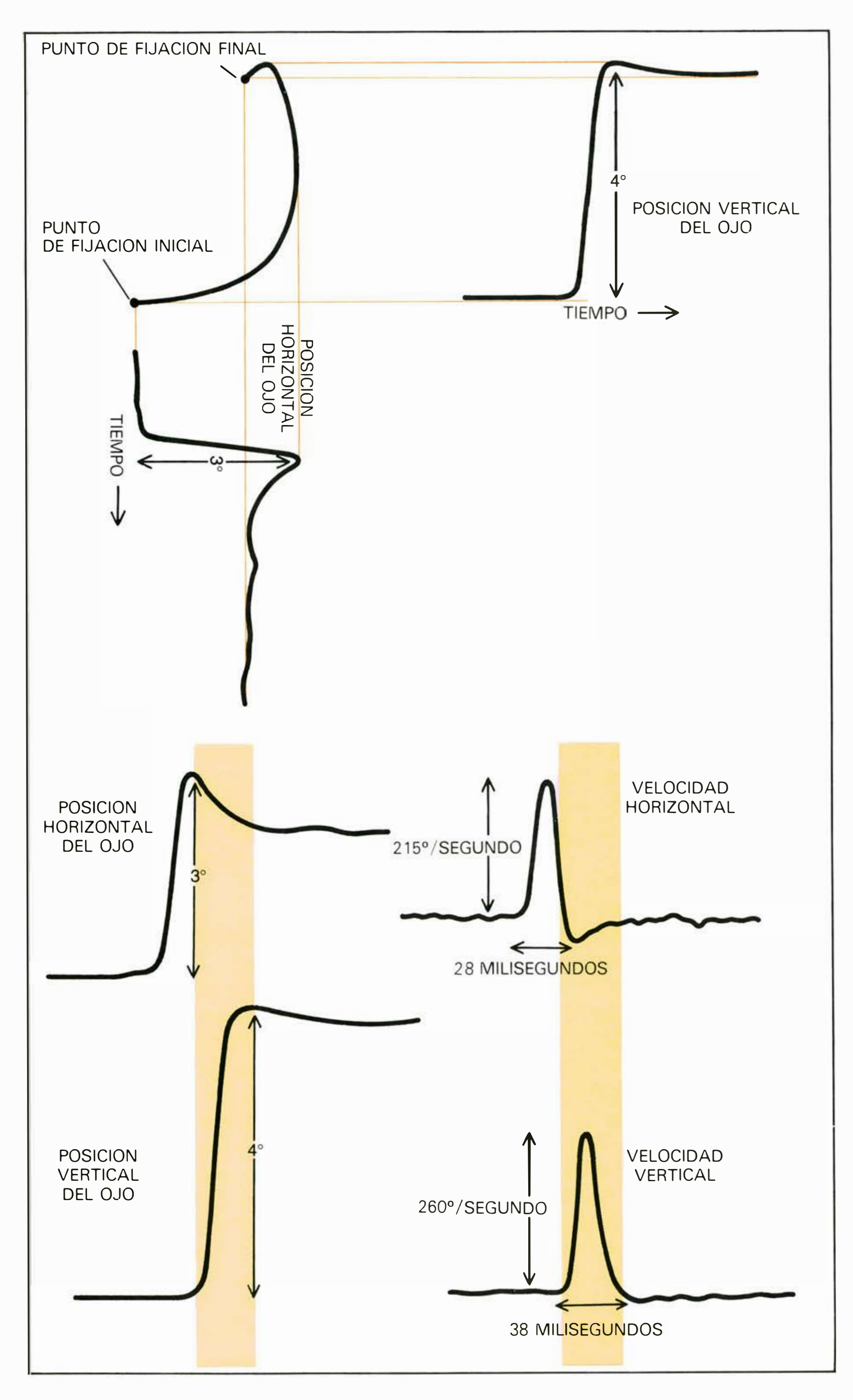
Hay que destacar que las sacudidas de mayor magnitud cansan más que las más cortas. Tras la realización de tan sólo 30 sacudidas de 50 grados u 80 sacudidas de 30 grados, un sujeto típico no puede efectuar ya sacudidas de forma normal. Además, si un sujeto llega a fatigarse con saltos del blanco cortos, sigue mostrando señales de fatiga si se le expone a saltos del blanco mayores, mientras que un sujeto fatigado por saltos del blanco largos en ocasiones puede efectuar sacudidas menores normales antes de que el efecto de la fatiga se deje notar nuevamente.

Utilizamos la palabra fatiga en un sentido muy general, ya que en nuestros experimentos no existe modo alguno de diferenciar la fatiga muscular de la adaptación sensorial a los estímulos ni de la habituación a los impulsos a nivel de sistema nervioso central. En cualquier caso, la fatiga hace surgir desviaciones en las señales de control neurológico, creando así variaciones en las trayectorias de las sacudidas. Las sacudidas solapadas quizá sean las más irregulares de todas las variaciones de las sacudidas. Las simulaciones realizadas en el modelo de inervación recíproca sugieren que las sacudidas solapadas están compuestas por dos sacudidas tan próximas una de otra que el globo ocular no tiene tiempo para reducir su velocidad a cero antes de que se inicie la segunda sacudida. Sin embargo, ello no significa que las porciones correspondientes a las descargas de las dos señales de control se hayan de solapar; la descarga de una sacudida normal suele finalizarse a medio camino en el movimiento de sacudida, de tal manera que en el par de sacudidas próximas existe una breve caída en la actividad de elevada frecuencia de las motoneuronas, caída que ocurre entre ambas sacudidas. Los parámetros de secuencia principal de estos pares de sacudidas indican que la generación de estos movimientos no se explica simplemente por una interrupción en la señal neurológica de una sacudida que hubiese recorrido la misma distancia que las dos pequeñas sacudidas solapadas. Si las sacudidas individuales fueran sencillamente una descomposición temporal de la sacudida conjunta, la suma de sus parámetros –aunque no los propios parámetros individuales– se ajustaría a los datos de la secuencia principal. Pero resulta que no es así. Por el contrario, los parámetros individuales son los que se adaptan a la secuencia principal, lo cual implica que lo que se ha dividido en dos es la magnitud de la sacudida conjunta y no la duración de la misma. Por consiguiente, cada una de las sacudidas componentes de un par de solapamiento se genera por una señal normal –tipo descarga y cambio de ritmo– para una sacudida de su magnitud. Puesto que las sacudidas solapadas no son el resultado sencillo de una pausa en la actividad de las motoneuronas, el centro del control motor que construye las señales de control neurológico tiene que decidirse con antelación a abandonar la sacudida única y la estrategia del control temporalmente óptimo, pasando a construir un par de sacudidas solapadas, cada una de las cuales es temporalmente óptima para su magnitud.

Pasemos a analizar ahora las sacudidas oblicuas, cuyo movimiento presenta una componente vertical y una componente horizontal. El análisis de la secuencia principal de estas componentes pone de manifiesto que, en ambas direcciones, el movimiento es básicamente del tipo sacudida. Las sacudidas verticales son algo más lentas que las horizontales, de modo que, en el diagrama de secuencia principal, se hallan ligeramente corridas respecto a los datos de las sacudidas horizontales. En cuanto a los demás aspectos se refiere, las sacudidas verticales y horizontales son semejantes en su estructura dinámica conjunta. Además, son los mismos tipos de irregularidades los que aparecen tanto en las sacudidas verticales como en las horizontales, y con la misma frecuencia. Ya que no parece existir ninguna diferencia intrínseca entre las trayectorias de ambas componentes, no hay necesidad

alguna de plantear aquí los orígenes neurológicos de las variaciones en las sacudidas verticales: sobrepasadas dinámicas, sacudidas solapadas, etcétera. Sin embargo, aún queda por contestar una cuestión de gran interés: ¿qué relación existe entre las trayectorias de ambas componentes de los movimientos oculares de sacudida oblicua?

Las componentes horizontal y vertical de las sacudidas oblicuas son fisiológicamente independientes, es decir, se llevan a cabo mediante la participación de vías nerviosas separadas con centros de control motor distintos, diferentes motoneuronas y músculos oculares también diferentes. Creemos, además, que las componentes son temporal y dinámicamente independientes, con trayectorias de sacudida no relacionadas entre sí sino generadas por sistemas distintos de señales de control neurológico. Existen dos observaciones que abogan a favor de esta tesis: la primera es que no parece existir relación alguna entre la duración y situación en el tiempo de cada una de las componentes; la segunda es que las variaciones concretas que aparecen en la trayectoria de una de las componentes no parecen tener relación tampoco con las variaciones concretas que aparecen en la trayectoria de la otra.



MOVIMIENTO OCULAR DE SACUDIDA OBLICUA (*arriba*). Tiene una componente de movimiento vertical además de la componente horizontal; cada una de estas componentes constituye una sacudida. Como nos indican las representaciones gráficas de posición y velocidad de ambas componentes (*arriba*) de las sacudidas oblicuas, parece que dichas componentes son dinámicamente independientes, de duraciones distintas y trayectorias no sincronizadas. En este caso, la sacudida horizontal presenta sobrepasada dinámica, pero no así la sacudida vertical; además, la sacudida horizontal está casi acabando cuando empieza la vertical (*franja de color*). Ambas componentes son independientes.

La independencia temporal de las componentes se demuestra porque las sacudidas oblicuas suelen ser raras veces rectas. Obviamente, una sacudida sólo puede ser recta si la componente horizontal y la vertical inician y concluyen su movimiento simultáneamente. El análisis de la secuencia principal de las componentes de las sacudidas oblicuas típicas demuestra que, en la mayoría de los casos, las componentes de la sacudida tienen duración diferente. Por ejemplo, una sacudida típica, de cinco grados de magnitud e inclinada de modo que forme un ángulo de 53 grados con la horizontal, presenta una componente horizontal de tres grados que dura 30 milisegundos y una componente vertical de cuatro grados que dura 40 milisegundos, de modo que resulta claramente imposible que ambas componentes inicien y concluyan su movimiento simultáneamente. La componente más corta de una sacudida oblicua puede iniciarse en cualquier momento del transcurso de la componente más larga, dándose incluso el caso de que se concluya una componente antes de que se inicie la otra, formándose una trayectoria en forma de *L*.

La independencia dinámica de las vías nerviosas horizontal y vertical se demuestra con la curvatura que presentan las trayectorias oblicuas. En otras palabras, las irregularidades en la forma de una sacudida oblicua son resultado de las diferencias entre las estructuras dinámicas de ambas componentes. Así no parece que exista relación entre la aparición de una sobrepasada dinámica en la trayectoria de una componente y su aparición en la trayectoria de la otra.

Como ya hemos demostrado, las variaciones en las trayectorias de las sacudidas se originan por diversos tipos de irregularidades en las señales de control compuestas por descarga y cambio de

ritmo: inversiones en la actividad nerviosa de las motoneuronas, desajustes en la descarga o en el cambio de ritmo, etcétera. Por tanto, el hecho de la independencia dinámica de las trayectorias de ambas componentes es una prueba más que confirma nuestra creencia de que las trayectorias son generadas por señales de control originadas a su vez en centros de control motor distintos, situados en diferentes zonas del cerebro. Hemos observado que las trayectorias de las sacudidas oblicuas pueden presentar formas mucho más variadas que las de las sacudidas horizontales, lo cual era de esperar al tratarse de movimientos en que se pueden introducir irregularidades a través de desviaciones en dos grupos distintos de señales neurológicas.

Hay que apuntar, sin embargo, que en realidad sí existe cierto intercambio de información —quizá tan sólo ruido de fondo— entre las vías nerviosas a través de las cuales se llevan a cabo los movimientos oculares de sacudida horizontales y verticales. Por ejemplo, si se somete un sujeto a saltos del blanco exclusivamente horizontales, es raro que se originen sacudidas totalmente lineales. En la mayoría de los casos, la sacudida horizontal creada por un estímulo del blanco del tipo mencionado se ve acompañada por un movimiento vertical pasajero, observación que fue realizada por Raymond Dodge, de la Universidad de Wesleyan, hace unos 80 años. Al parecer, el ruido procedente de la vía del movimiento horizontal afecta también a la vía del movimiento vertical y provoca la aparición de una componente vertical extraña.

Por tanto, cuando además de las componentes horizontales de los movimientos oculares se consideran las verticales, se observa una mayor variabilidad en las trayectorias de las sacudidas. Los movimientos oculares de sacudida repetidos entre dos puntos fijos suelen tener trayectorias diferentes y a menudo son muy diferentes. Aunque las sacudidas oblicuas repetidas tengan componentes horizontales y verticales idénticas, casi siempre difieren de un movimiento a otro las relaciones dinámicas y temporales de ambas componentes (y, por tanto, también difieren las relaciones entre las señales de control neurológico que se hallan tras cada par de componentes). En el sistema del movimiento ocular, si ampliamos el análisis considerando a ambos ojos, encontramos una variabilidad todavía mayor. No hay que olvidar que las sobrepasadas dinámicas, los deslizamientos, las sacudidas dobles y las solapadas son fenómenos monoculares.

A pesar de ello, resulta sorprendente que la variabilidad de las trayectorias de sacudida no parezca constituir una cuestión de importancia para el cerebro. La visión no se ve alterada por este hecho, quizá porque se suprime la visión inmediatamente antes y durante una sacudida. (Se puede demostrar que se da esta supresión desplazando la mirada de un lado a otro de una habitación. Aunque la imagen que existe en la retina se mueve claramente, no se percibe a la habitación como sujeta a un movimiento giratorio.) Además, anteriormente a cada sacudida, el cerebro hace una estimación del cambio esperado en el marco de referencia del ojo, de modo que el observador no es consciente, a continuación, del cambio habido en la imagen retiniana. En los casos en que los movimientos oculares de sacudida adquirieran caracteres extremos, parece probable que el cerebro intervendría modulándolos, ejerciendo algún tipo de control o recalibrado (posiblemente a nivel del cerebelo), para restarles rareza a las trayectorias.

En nuestra labor de investigación acerca de las trayectorias de los movimientos oculares de sacudida, hemos hecho un uso extensivo de las técnicas empleadas en bioingeniería, como lo demuestra nuestro aparato experimental, que fue expresamente diseñado para ajustarse a la elevada velocidad del ojo; la utilización de un ordenador digital para la recogida de datos, su posterior análisis y para las simulaciones; la aplicación del concepto de secuencia principal con el fin de encuadrar los diversos datos relativos a los movimientos oculares; y el desarrollo de un modelo de inervación recíproca como medio de conceptualización del control de los movimientos oculares. Estos dispositivos nos han permitido llegar a un conocimiento más profundo de la naturaleza y del control de varios tipos de movimientos oculares. Efectivamente, hemos logrado identificar el nivel de actividad del control neurológico en el cual se originan muchos tipos específicos de variaciones del movimiento ocular. Como consecuencia de este estudio, en muchos casos hemos conseguido demostrar la manera exacta en que la fatiga, la tensión y la enfermedad afectan a las señales de control neurológico creadoras de los movimientos oculares y, por tanto, a la función normal de este sistema. Actualmente se están desarrollando aplicaciones clínicas de estos descubrimientos y creemos que ayudarán a mejorar el diagnóstico y tratamiento de los pacientes que sufran diversos trastornos físicos.

Vibraciones acopladas de las cuerdas del piano

La mayoría de las notas de un piano se producen por la vibración simultánea de dos o tres cuerdas. Estas cuerdas no están exactamente afinadas a la misma frecuencia, lo que contribuye de forma inesperada al sonido del instrumento

Gabriel Weinreich

En 1709 el constructor de clavicordios Bartolommeo Cristofori construyó el prototipo del piano moderno, en el que unos martillos percuten las cuerdas. Desde entonces, los pianos se han construido siempre con más de una cuerda por nota. El teclado del moderno Steinway Modelo B tiene 88 teclas, donde las 68 más agudas actúan sobre tres cuerdas y las 20 más graves actúan sobre dos cuerdas o una sola. Si bien el hecho de utilizar dos o tres cuerdas fue introducido ya a mediados del siglo XVII para aumentar la potencia sonora, recientes estudios acústicos han demostrado que en el piano, por lo menos, la utilización de grupos de dos y tres cuerdas afecta también a la calidad del sonido. De estos estudios se deduce que, al percutir por separado cada cuerda de un grupo formado por dos o tres, sus frecuencias difieren ligeramente, en apariencia de una forma casual, incluso después de su afinado por los mejores especialistas en pianos. Los oyentes prefieren el sonido producido por una tecla, cuando existe una pequeña diferencia entre las frecuencias propias, al sonido producido por una tecla cuando no existen tales diferencias.

Mis investigaciones en la Universidad de Michigan se han centrado en cómo estas pequeñas "desafinaciones", o diferencias de frecuencia, contribuyen a la sonoridad del piano. También he estudiado otras posibilidades de modificación de sonoridad debidas a las cuerdas dobles o triples con respecto a lo que hubiera sido si los pianos se hubieran construido con una sola cuerda por nota. De acuerdo con esos estudios y con los de otros investigadores, puede apreciarse que el sonido primero se atenúa rápidamente (el "sonido inmediato") y luego lo hace con lentitud (la "resonancia").

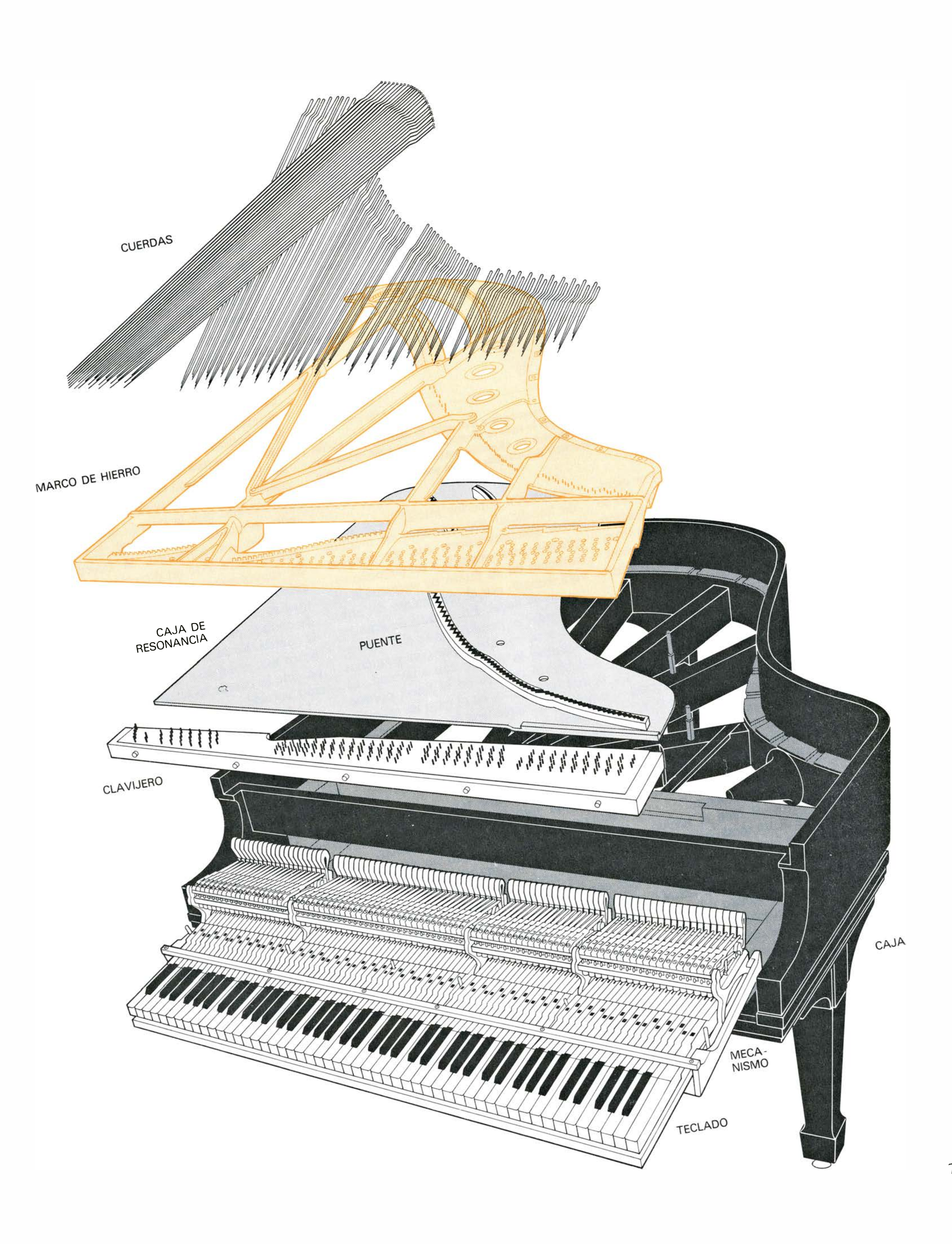
Al pulsar una tecla del piano, un martillo recubierto de fieltro percute el grupo unísono correspondiente de dos o tres cuerdas. Al mismo tiempo, el bloque de fieltro blando, llamado apagador, se levanta alejándose de las cuerdas para que éstas puedan vibrar libremente. Al soltar la tecla, entra en acción el apagador que hace que las cuerdas dejen de vibrar. También el sonido cesará si se deja que las vibraciones se extingan por sí mismas hasta su inaudibilidad.

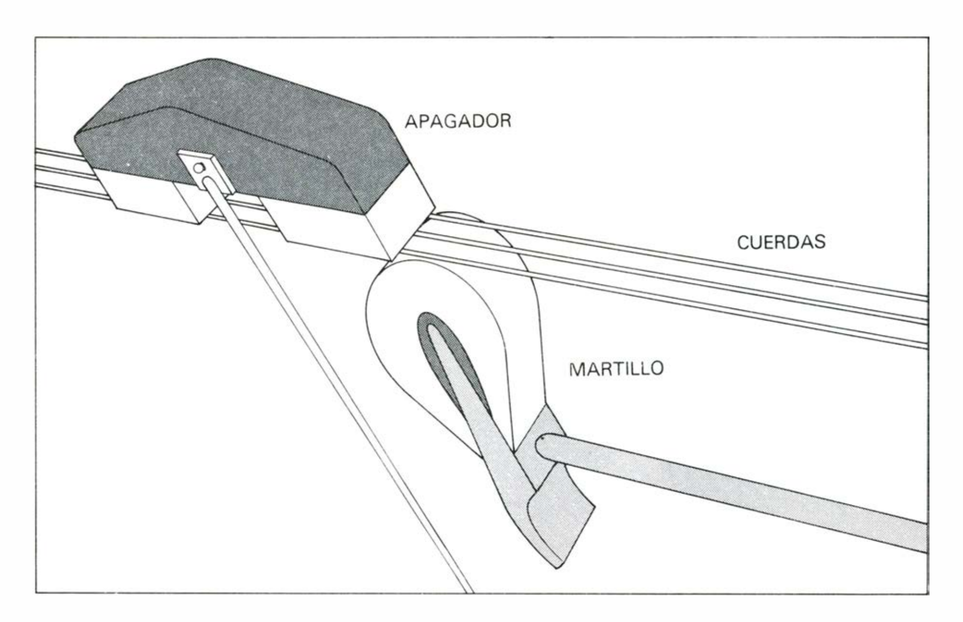
El sonido musical se produce gracias a las vibraciones internas del instrumento. En los instrumentos de viento, como la flauta, y los instrumentos de cuerdas tocados con arco, como el violín, las vibraciones son "sostenidas" por la aportación de energía en el transcurso de cada oscilación; en el piano, por contra, las vibraciones son "libres": no se aporta más energía después del impacto inicial del martillo. Por consiguiente, las características musicales del piano dependen principalmente de cómo se disipa la energía de las cuerdas.

La energía se disipa a través de las fuerzas de rozamiento, que se manifiestan en diferentes secciones del piano. Cuando una cuerda vibra con amplitudes pequeñas, la disipación de la energía es "lineal": la velocidad de pérdida de energía de la cuerda es proporcional a la cantidad de energía que tiene almacenada. En un sistema lineal de este tipo, la disminución de la vibración es exponencial: fracciones de energía iguales se disipan en intervalos de tiempo iguales. La desintegración radiactiva del carbono 14 constituye un ejemplo típico de un sistema lineal análogo a la disminución de la vibración de una cuerda de un piano. La velocidad a la que un material pierde carbono 14 es proporcional a la cantidad de carbono 14 que contiene. El que se desintegre la mitad del carbono 14 cada 5730 años significa que el ritmo de desintegración es exponencial, como ocurre con la energía de una cuerda del piano.

La relación entre las amplitudes de presión de dos sonidos se mide en decibelios; cada 20 decibelios corresponden a una variación en la presión acústica de un factor 10. Por tanto, una reducción de 40 decibelios corresponde a una reducción en la presión acústica de dos factores 10, esto es, un factor 100; una reducción de 10 decibelios corresponde a una reducción en la presión acústica de un factor de $\sqrt{10}$, o sea, 3,16. Cuando la presión acústica se especifica por medio de su valor relativo (en decibelios) con respecto a un valor de referencia se denomina nivel de la presión acústica. Una ventaja de este enfoque consiste en que una disminución exponencial corresponde a una línea recta en una gráfica del nivel de presión acústica en función del tiempo. La presión acústica es una medida física similar a la intensidad, aunque no debe confundirse con ella. La presión acústica es un fenómeno estrictamente físico mientras que la intensidad es un fenómeno psicofísico que no depende sólo de las propiedades físicas del sonido, sino de cómo responden ante

DESPLIEGUE ILUSTRATIVO de un piano Steinway modelo B, en la página opuesta, donde se muestran los principales componentes. El teclado tiene 88 teclas que cubren 7 octavas y un tercio. Las 68 teclas más agudas *(derecha)* ponen en movimiento grupos de tres cuerdas y las 20 más graves *(izquierda)* ponen en movimiento pares de cuerdas o cuerdas solas. Conectado al teclado se halla el mecanismo que comprende martillos y apagadores, que determinan los movimientos de las cuerdas. Al pulsar una tecla, el martillo hace vibrar las cuerdas. Las cuerdas cruzan el puente de madera que transmite las vibraciones a la caja de resonancia, desde la cual las vibraciones se irradian hacia el aire. También se irradia sonido, en menor grado, desde otras secciones del piano. (Dibujo de Dan Todd.)


CUERDAS
MARCO DE HIERRO
CAJA DE
RESONANCIA
PUENTE
CLAVIJERO
CAJA
MECA-
NISMO
TECLADO



MARTILLO PERCUTOR DE LAS CUERDAS que corresponden a una nota. El martillo percute las cuerdas con la misma fuerza y al mismo tiempo. Dado que lo hace en dirección vertical éstas se mueven en gran parte en dicha dirección. También se mueven algo en dirección horizontal. Este movimiento podría venir causado por irregularidades en la superficie del martillo o en la posición de las cuerdas.

el mismo el oído y el cerebro. Por ejemplo, al doblar la presión acústica, no se percibe necesariamente una duplicación de la intensidad. Sin embargo, para fines cuantitativos, basta con tratar la presión acústica y la intensidad como si fueran una misma cosa.

La atenuación del sonido de una cuerda de piano (cuando no se permite que vibren las restantes cuerdas de su propio grupo) es realmente mucho más complicada que lo que sugiere una línea recta. La gráfica del nivel de presión acústica en función del tiempo no resulta ser una línea recta sino una línea quebrada formada por dos segmentos rectilíneos, el primero empezando a un nivel alto y atenuándose con rapidez y el segundo, a continuación, a un nivel inferior y atenuándose lentamente.

Esto significa que, si bien el sonido se atenúa siempre de un modo exponencial, el gradiente de la atenuación exponencial cambia de una forma brusca en un punto. La rápida atenuación inicial caracteriza el sonido inmediato y la atenuación final lenta caracteriza la resonancia. La presencia de estos dos tipos de sonido es un rasgo peculiar de la sonoridad del piano. El sonido inmediato es una especie de "ping" similar al de un xilofón, pero mientras éste enmudece al cabo de unos segundos, el piano continúa sonando. Este efecto constituye la resonancia, que permite sostener melodías con el piano, cosa imposible de realizar con el xilofón.

La presencia de sonido inmediato y resonancia no es resultado de una falta de linealidad: fuerzas de rozamiento proporcionalmente mayores a presiones acústicas superiores son la causa de una atenuación superior inicial. Tales faltas de linealidad son comunes en muchos otros sistemas físicos. Así, un péndulo que sólo esté sometido a un pequeño rozamiento en su punto de suspensión se verá frenado por el rozamiento del aire, que no es lineal. En el caso de grandes amplitudes el péndulo agita con violencia el aire, mientras que con pequeñas amplitudes lo mueve con suavidad. El rozamiento del aire no es lineal porque el aire turbulento, en el caso de grandes amplitudes, ejerce una fuerza proporcionalmente superior sobre el péndulo que cuando el aire se mueve suavemente con amplitudes pequeñas.

Puede demostrarse por vía experimental que la resonancia de un piano no se debe a un fenómeno similar. Si las fuerzas de rozamiento no fueran lineales, el paso de una atenuación rápida a una atenuación lenta tendría lugar con la misma amplitud, con independencia del desplazamiento inicial de la cuerda. El desplazamiento inicial de la cuerda afectaría sólo a la duración del sonido inmediato: cuanto mayor sea el desplazamiento mayor será la duración del sonido inmediato. Ello significa que las gráficas del nivel de presión acústica en función del tiempo, realizadas para diferentes desplazamientos iniciales, darían traslaciones horizontales de una con respecto a la otra.

El cambio de un atenuación rápida a una atenuación lenta ocurre realmente a diferentes presiones acústicas, aunque después del mismo número de segundos tras el desplazamiento inicial. En otras palabras, el sonido inmediato tiene la misma duración con independencia de la magnitud del desplazamiento inicial de la cuerda. Las gráficas de las presiones acústicas son simples traslaciones verticales una de otra, que no hay que confundir con las traslaciones horizontales indicativas de un rozamiento no lineal. En un sistema lineal, como el de una cuerda de piano, la amplitud del movimiento puede incrementarse o disminuirse uniformemente, sin cambiar el tipo de movimiento.

Pero, ¿cuál es la causa del cambio del gradiente de atenuación? Para una sola cuerda, el cambio se debe a la existencia de dos polarizaciones de vibración en la cuerda. Me referiré a ellos como vertical y horizontal, de acuerdo con las direcciones reales en un piano de cola. Otros movimientos, incluidos los circulares y los elípticos, pueden considerarse como superposiciones de las dos polarizaciones básicas. Como el martillo percute la cuerda en dirección vertical, puede parecer extraño que la cuerda adquiera algún movimiento horizontal. En realidad, en un comienzo, la cuerda se mueve sobre todo en dirección vertical. El movimiento horizontal inicial que existiese pudiera ser el resultado de pequeñas irregularidades en la superficie de choque del martillo o la posición de la cuerda. Es decir, el ligero movimiento horizontal obedece a que el martillo no percute la cuerda exactamente en dirección vertical.

Utilicé un detector electrónico sensible para medir separadamente los movimientos verticales y los horizontales de una cuerda. Mis mediciones mostraron que cada polarización por separado se atenúa de un modo exponencial, si bien la polarización vertical se atenúa con mayor rapidez [*véase la ilustración superior de la página siguiente*]. Esto significa que, aunque el movimiento vertical es mucho más potente que el movimiento horizontal (probablemente la relación sea superior a 10), llega un momento en que domina el movimiento horizontal. Ello provoca que el movimiento vertical dé lugar a un sonido inmediato, de corta duración, y el movimiento horizontal dé lugar a una resonancia prolongada.

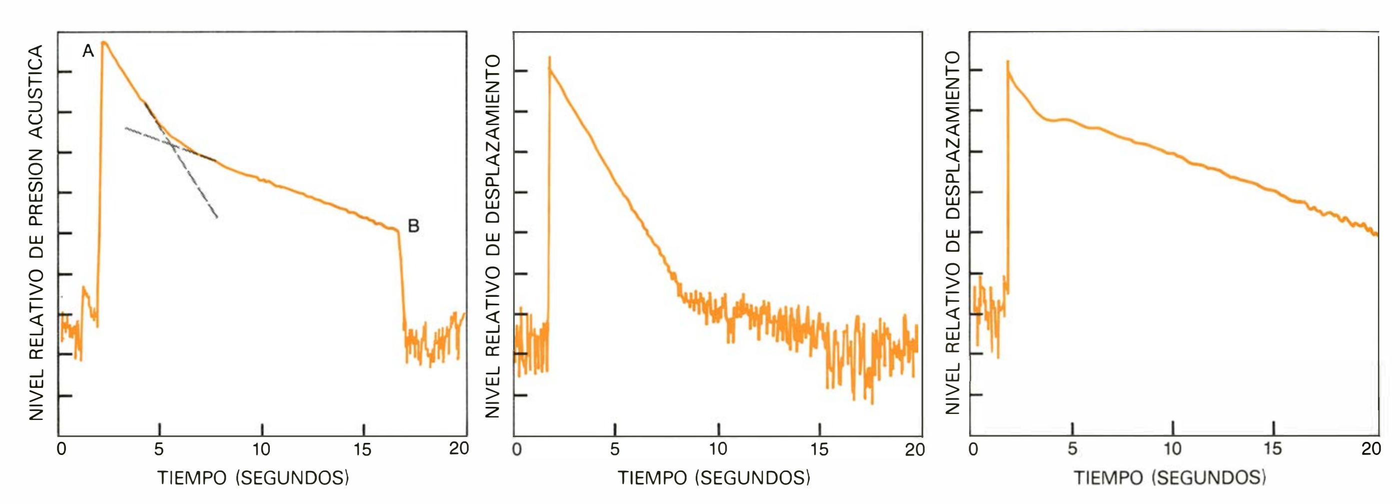
¿Por qué varía la atenuación en los dos tipos de movimiento? La respuesta depende de las distintas maneras como la cuerda puede disipar la energía. Cuando la cuerda pierde energía vibratoria, ésta puede emplearse en calentar la misma cuerda (rozamiento interior), en mover el aire adyacente (viscosidad y radiación sonora) o perderse por los apoyos de la cuerda. En el caso del piano predomina el último mecanismo. El extremo de la cuerda, hacia la parte del teclado, está fijo a un marco de hie-

rro, mientras que el extremo opuesto pasa por un "puente" de madera que se halla pegado a la caja de resonancia. El puente no es totalmente rígido, pues tiene por misión que la caja de resonancia vibre sincrónicamente con la cuerda. La caja de resonancia irradia al aire la mayor parte del sonido y, en menor grado, lo irradian también otras partes del piano. Ocurre que el movimiento vertical de la caja de resonancia "cede" más que el movimiento horizontal, y así la energía es fácilmente transferida del movimiento vertical de la cuerda al movimiento vertical de la caja de resonancia. Ello justifica la rápida atenuación del movimiento vertical de la cuerda, que es responsable del sonido inmediato. A su debido tiempo modificaré este esquema simplista, porque la mera "cesión" no conlleva necesariamente una pérdida de energía.

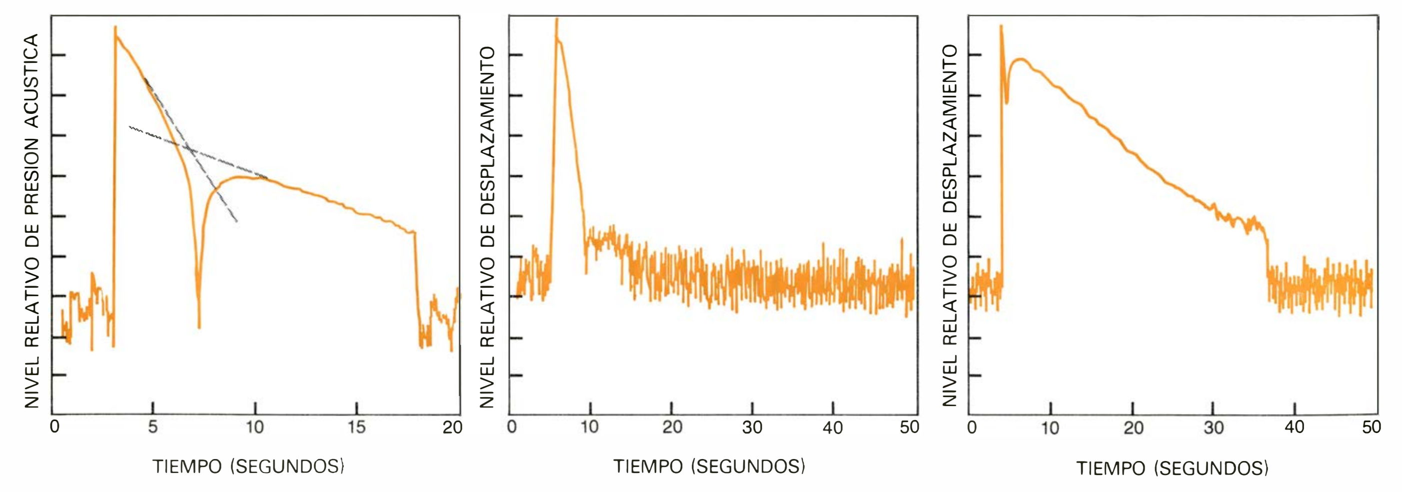
Los expertos en física acústica no han investigado aún con suficiente detalle cómo irradian las vibraciones horizontales. Realicé un experimento sencillo con el fin de observar si las ondas acústicas horizontales y las verticales salían emitidas por la misma fuente en el piano. (Hablando en sentido estricto, las ondas sonoras no son horizontales ni verticales. Estos términos se refieren aquí a las polarizaciones de las cuerdas correspondientes a las ondas sonoras.) Si se emiten a partir de fuentes diferentes, si son de una fuerza equiparable, se interfieren o se combinan para formar una onda sonora cuya amplitud, en cualquier punto dado de la sala, es igual a la suma de las amplitudes de las ondas componentes.

Donde las ondas se hallen exactamente en fase (las crestas coincidan y los valles también), aumentará la amplitud, y donde estén desfasadas (crestas coincidiendo con valles) disminuirá la amplitud. Con el fin de averiguar si era detectable este fenómeno de interferencia, coloqué un micrófono en distintos puntos del piano y examiné el comportamiento del nivel de presión acústica durante la transición del sonido inmediato a la resonancia. Efectivamente, en el momento de la transición encontré los puntos donde las ondas componentes se reforzaban y los puntos donde se destruían entre sí. Esto significa que la fase de las ondas componentes es distinta en diferentes puntos, confirmando que las ondas horizontales y verticales se emiten por "antenas" distintas.

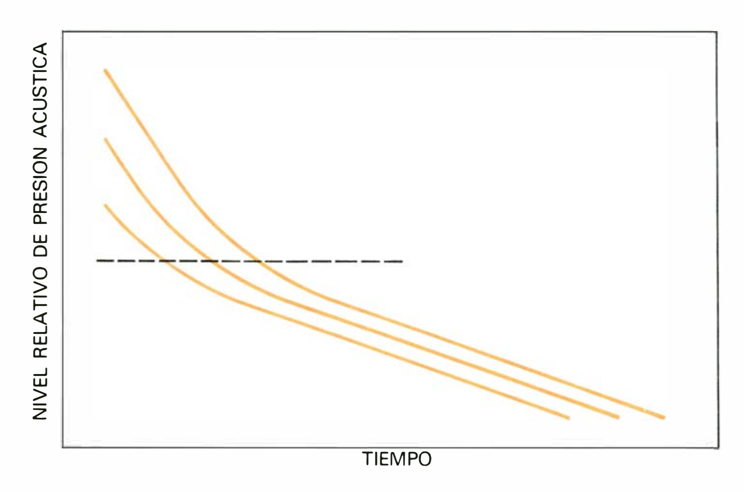
La presencia de movimiento horizontal constituye sólo uno de los factores que producen la resonancia, un factor que no depende de que las cuerdas se presenten formando grupos de

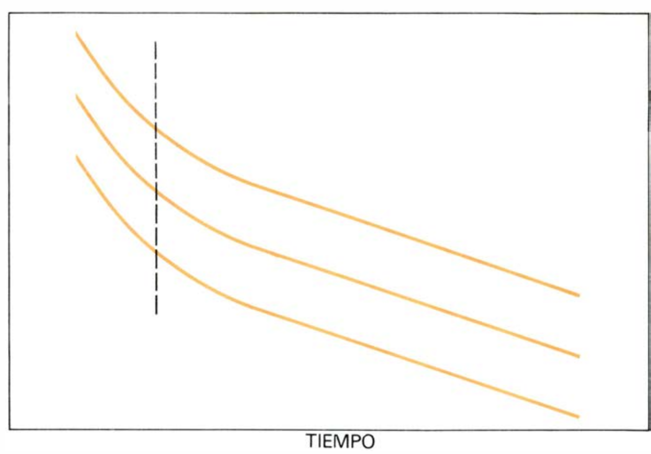


NIVEL RELATIVO DE PRESION ACUSTICA, medido en decibelios, presentado en forma de gráfica en función del tiempo para una sola cuerda vibrante del piano *(izquierda)*. Cada división de la escala vertical corresponde a 10 decibelios. En el instante *A* se pulsa la tecla y el martillo percute las cuerdas; en el instante *B*, se suelta la tecla y el apagador detiene los movimientos de las cuerdas. La gráfica se divide en dos segmentos rectilíneos, donde las líneas a trazos constituyen sus extensiones. La atenuación inicial, rápida y de poca duración, forma el sonido inmediato y la atenuación final, larga y lenta, que da al timbre del piano su calidad cantabile, constituye la resonancia. El ruido que se detecta en el instrumento electrónico de registro determina que se observen curvas irregulares antes de *A* y después de *B*. Cuando sólo vibra una cuerda de un grupo, el movimiento vertical de la cuerda *(centro)* da lugar al sonido inmediato, y el movimiento horizontal de la cuerda *(derecha)* origina la resonancia. (Ilustración de Dan Todd.)

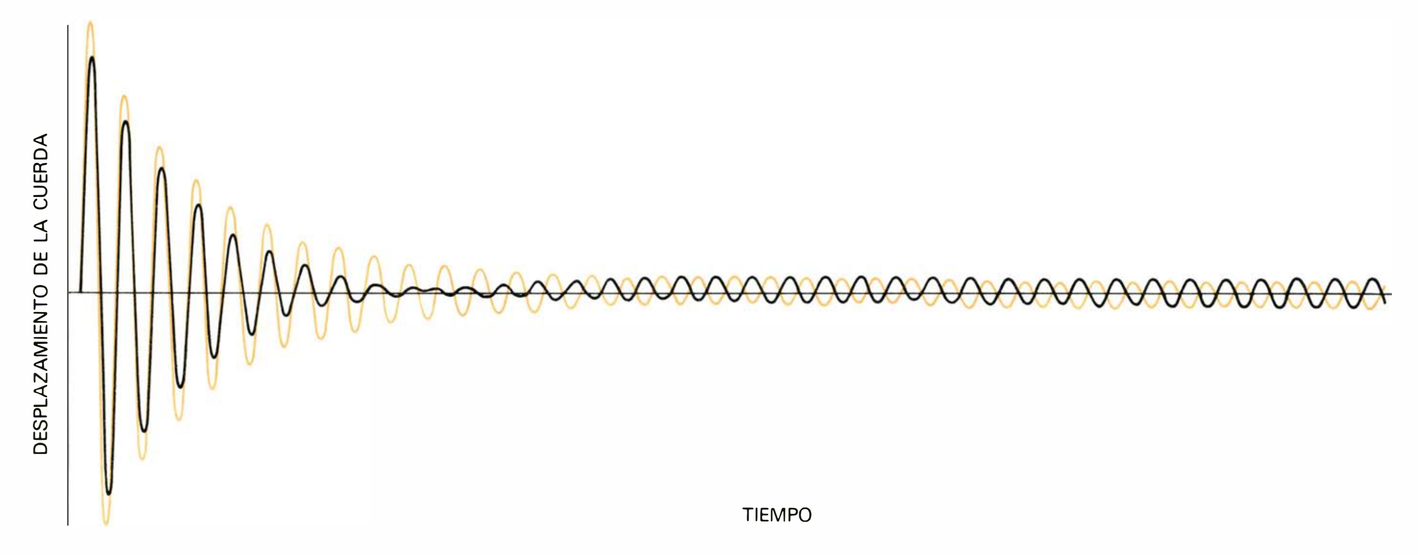


CURVAS DE INTERFERENCIA entre ondas sonoras correspondientes a las vibraciones verticales y horizontales de una cuerda. Dicha configuración muestra que tales ondas emanan de fuentes diferentes en el piano *(izquierda)*. La gráfica de la presión acústica relativa se hizo bajo las mismas condiciones que en la gráfica de la izquierda en la figura superior de esta página, con la salvedad de que se cambió la posición del micrófono de captación. En este punto, los sonidos debidos a vibraciones verticales y horizontales llegan en oposición de fase, de modo que se anulan cuando sus amplitudes son iguales. Al vibrar dos cuerdas en un grupo unísono *(derecha)*, el movimiento dura más que cuando sólo vibra una de ellas *(centro)*.

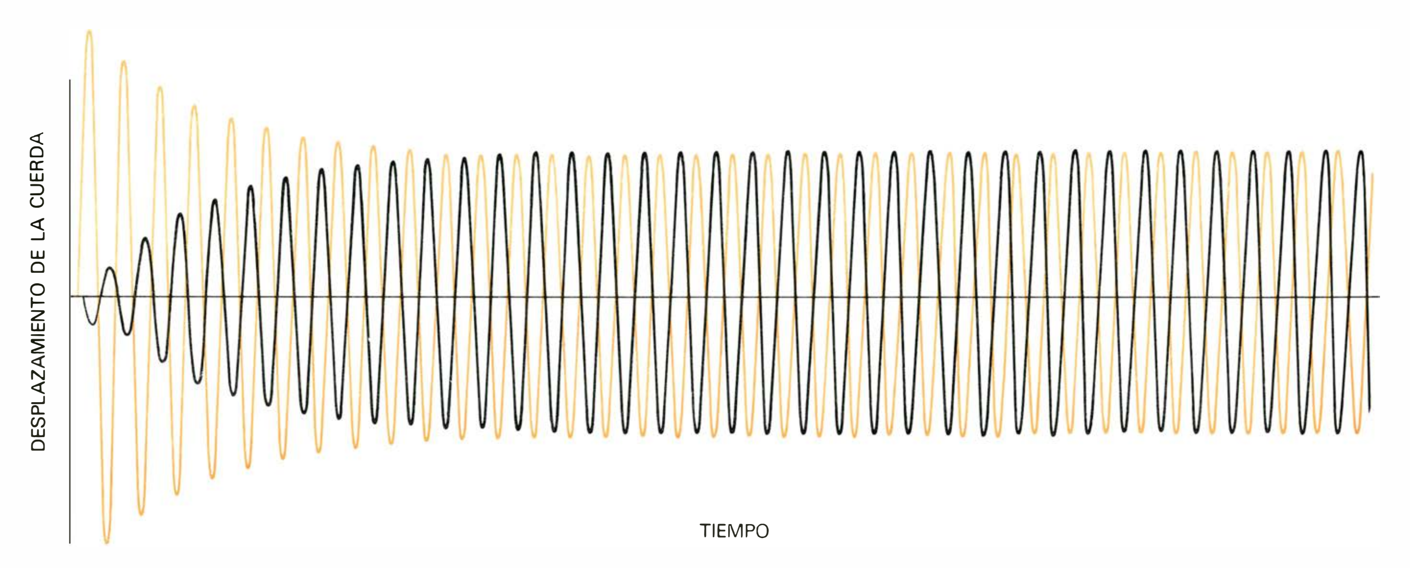




FUERZAS DE ROZAMIENTO, que son proporcionalmente mayores a presiones acústicas superiores; no hay que atribuirles la transición del sonido inmediato a la resonancia. Tales fuerzas no lineales de rozamiento provocarían que el cambio se produjera a las mismas presiones acústicas, independientemente del desplazamiento inicial de la cuerda. En ese caso, las gráficas del nivel de presión acústica en función del tiempo, para diferentes desplazamientos iniciales, se traducirían en traslaciones horizontales una de otra *(izquierda)*. Las gráficas son translaciones verticales *(derecha)*, lo que alude a un rozamiento lineal. El sonido inmediato tiene la misma duración, con independencia del desplazamiento inicial de la cuerda.



IMPERFECCIONES ATRIBUIBLES AL MARTILLO. Pueden tener como resultado amplitudes en las cuerdas que no sean absolutamente iguales. Se ponen en movimiento dos cuerdas a la vez, pero presentando la cuerda coloreada una amplitud mayor que la cuerda negra. Los movimientos de las cuerdas empiezan a atenuarse, y cuando la amplitud de la cuerda negra se aproxima a cero, el puente continúa moviéndose porque todavía está forzado a hacerlo por la cuerda coloreada. La cuerda negra no sólo alcanza la amplitud cero sino que la "rebasa", formando una vibración de fase opuesta a partir de la energía que absorbe del puente. Los movimientos guardan perfecta asimetría, movimiento que da lugar a la resonancia.



PEDAL UNA CORDA, o pedal suave. Aumenta la proporción de resonancia con respecto al sonido inmediato, desplazando todo el teclado de suerte que un martillo sólo percuta una cuerda de un par. La cuerda no percutida *(negro)* empieza a absorber energía del puente que vibra sincrónicamente con la otra cuerda *(en color)*. La cuerda no percutida comienza en seguida a moverse en oposición de fase con la otra cuerda. De ello se sigue un movimiento que es asimétrico desde el principio, lo cual permite que el timbre del piano conserve su calidad cantabile para los fragmentos tranquilos.

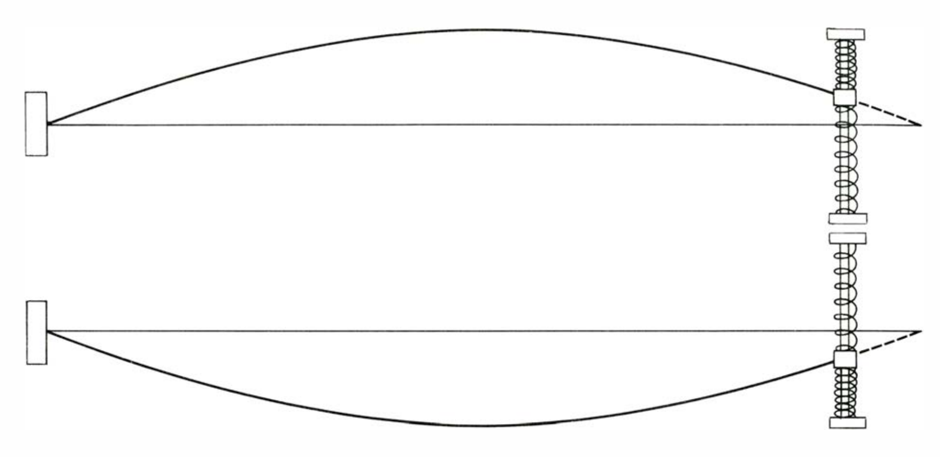
dos o tres. Por tanto, se seguiría percibiendo una resonancia debida al movimiento horizontal si los pianos se construyeran con una cuerda por nota. Sin embargo, el triplicar o doblar las cuerdas contribuye también a la resonancia. Las cuerdas que forman un grupo unísono cruzan el puente muy próximas entre sí. Así resulta que sus movimientos están acoplados: cuando una cuerda vibra, el puente vibra con ella y transmite este movimiento a las otras cuerdas del grupo unísono. La experimentación demuestra que el desplazamiento vertical de una de estas cuerdas vibrantes acopladas se atenúa a veces hasta una velocidad lenta, que es típica de la resonancia. Por consiguiente, el movimiento horizontal sin acoplamiento y el movimiento vertical con acoplamiento contribuyen, de un modo independiente, a la resonancia.

¿Por qué el movimiento acoplado rebaja la atenuación en vez de aumentarla o mantenerla? Para explicar este fenómeno hay que introducir un dato: que la velocidad a la que se disipa la energía a través del movimiento del puente no está determinada porque el puente pueda o no moverse, sino por cómo se mueve. En particular, si dos cuerdas cruzan el puente por el mismo punto y vibran a la misma frecuencia y amplitud, pero en oposición de fase (una cuerda sube cuando la otra cuerda está bajando), la fuerza resultante sobre el puente será cero. Y el puente no se moverá. Para cada cuerda, por separado, el puente parecerá completamente rígido. Pero si las cuerdas vibran en fase (incluso con la misma amplitud y frecuencia), el movimiento del puente será doble que si vibrase una sola cuerda. La atenuación, consiguientemente, será también doble. Los mismos principios que gobiernan la disipación de energía en dos cuerdas son válidos para tres cuerdas, aunque en el último caso la dificultad en seguir tres factores de fase puede ocultar lo que está pasando.

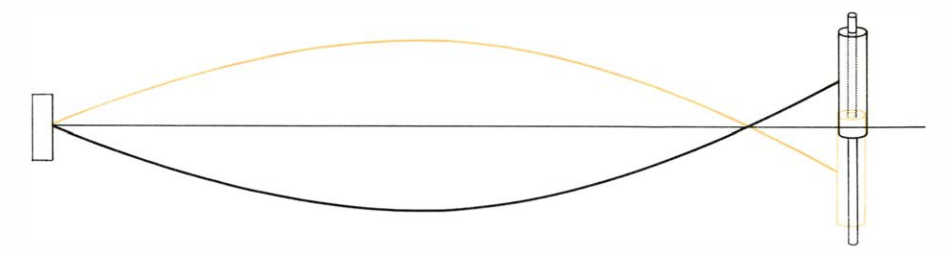
En la mayoría de las situaciones de que da cuenta la física acústica, los movimientos de dos cuerdas ni son exactamente iguales (simétricos) ni exactamente opuestos (asimétricos). En un piano, los movimientos de las cuerdas de un grupo unísono poseerán, en un comienzo, una simetría casi perfecta, ya que el martillo, en apariencia, percute las cuerdas con la misma fuerza y al mismo tiempo. Las pequeñas imperfecciones del martillo producirán amplitudes en la cuerda que no serán absolutamente iguales. Consideremos el caso de dos cuerdas que se muevan en fase, pero la primera cuerda lo haga con una amplitud mayor que la segunda. Inicialmente, ambas cuerdas perderán energía y cada una la perderá con mayor rapidez que si vibraran separadamente, ya que la otra "ayuda" a que el puente se mueva.

Cuando la amplitud de la segunda cuerda (aquélla con la amplitud menor) se acerca a cero, el puente continúa moviéndose porque es forzado por la primera. La segunda cuerda no sólo alcanza la amplitud cero sino que "se pasa", creando una vibración de fase opuesta a expensas de la energía del puente. Como las dos cuerdas se mueven ahora en fases opuestas, el movimiento del puente es menor que si una cuerda se moviera en ausencia de la otra. Las dos amplitudes se acercan asintóticamente en fases opuestas. Por último, los movimientos de las dos cuerdas son exactamente asimétricos. El movimiento inicial simétrico de las cuerdas constituye el sonido inmediato y el movimiento asimétrico de las cuerdas constituye la resonancia.

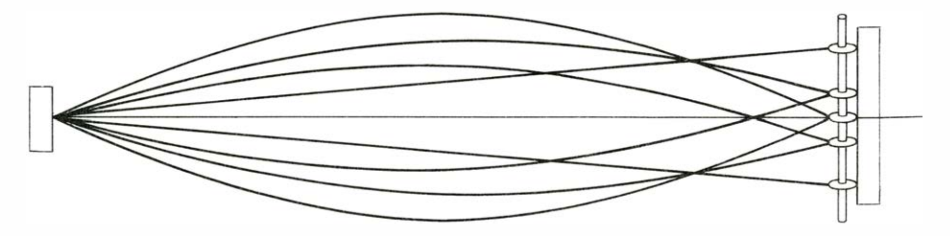
Una forma práctica de considerar la situación que acabo de describir consiste en imaginar el movimiento original como si fuera una superposición de dos movimientos: uno simétrico y otro asimétrico. Séame permitido ilustrar lo que



APOYO ELASTICO, que rebaja la frecuencia de una cuerda sin amortiguar su movimiento. Un ejemplo ideal de apoyo elástico lo constituye un anillo situado entre dos muelles, y que se desliza en ambos sentidos sin rozamiento sobre una varilla fija. El apoyo rebaja la frecuencia de la cuerda porque simula un movimiento que sería realizado por un trozo extra de cuerda: el anillo logra su desplazamiento máximo cuando lo alcanza la cuerda y consigue su desplazamiento cero cuando ésta lo consigue.



APOYO MACIZO, que aumenta la frecuencia de una cuerda sin amortiguar su movimiento. Tenemos un ejemplo ideal de un apoyo macizo en un gran bloque que pueda deslizarse sin fricción en ambos sentidos sobre una varilla fija. Aquí, el movimiento del bloque no depende de la fuerza de recuperación de los muelles, sino del efecto de la inercia. En las posiciones del bloque que se muestran, la cuerda debe tirar el bloque hacia atrás, contra la dirección de inercia, al objeto de modificar dicha orientación. El que la cuerda estire a menudo del bloque de gran masa significa que la cuerda se "cree" más corta de lo que mide en realidad. El apoyo ideal macizo incrementa, pues, la frecuencia de la cuerda.



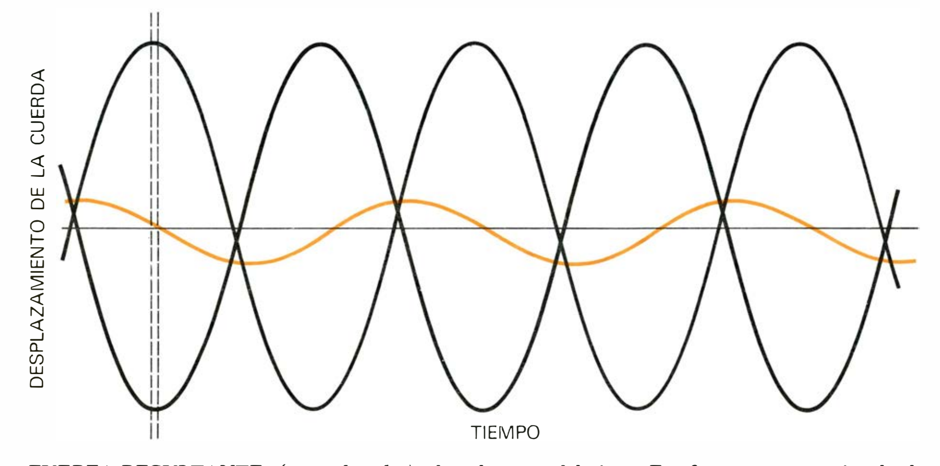
APOYO RESISTIVO, que no modifica la frecuencia de la cuerda, aunque sí amortigua su movimiento. Un anillo que se desliza en ambos sentidos sobre una varilla cuyo movimiento está retardado por el rozamiento contra la pared puede tomarse como ejemplo cabal de apoyo resistivo. El movimiento de la cuerda es amortiguado porque trabaja sobre el anillo para vencer el rozamiento entre el anillo y la pared. La diferencia de fase entre el desplazamiento del apoyo y la fuerza ejercida sobre el mismo es un cuarto de ciclo: cuando el anillo obtiene su desplazamiento máximo, la cuerda alcanza su desplazamiento cero; cuando ésta logra su desplazamiento máximo, el anillo alcanza su desplazamiento cero.

entiendo por superposición. Si tengo 10 manzanas y el lector tiene 6, podemos describir la situación como la superposición de un estado simétrico donde cada uno posee ocho manzanas y un estado asimétrico donde yo tengo dos manzanas y el lector menos dos manzanas. Esto, que es una simplicidad en el caso de las manzanas, resulta útil para explicar las vibraciones: nos permite pensar que la componente simétrica se atenúa a su velocidad característica y la asimétrica se atenúa a una velocidad menor (o en el caso ideal, no se atenúa). La suma algebraica de las dos amplitudes de la cuerda baja hacia cero, pero la diferencia permanece constante durante largo tiempo.

Puesto que el movimiento más general de dos cuerdas de un piano puede expresarse como superposición de un movimiento simétrico y un movimiento asimétrico, ambos tipos de movimiento son los modos normales del sistema de cuerdas del piano. Es interesante apuntar que el cambio en la atenuación de las vibraciones de las cuerdas del piano no constituye un fenómeno aislado en física. Las características de atenuación son, realmente, iguales a las de ciertas partículas subatómicas, como los kaones neutros. Dos tipos de kaón (K^0 y K^0), un par partícula-antipartícula, pueden formarse mediante la interacción fuerte en colisiones nucleares. Al estudiar las atenuaciones de los kaones, los físicos han identificado dos nuevas variedades de kaon: K^0_s, que se desintegra con rapidez, y K^0_l que lo hace lentamente. Resulta que K^0_s y K^0_l son, respectivamente, superposiciones simétricas y asimétricas de K^0 y K^0, al igual que el sonido inmediato y la resonancia son, respectivamente, las superposiciones simétricas y asimétricas de los movimientos de dos cuerdas de piano. En consecuencia, un haz que conste en un comienzo de sólo partículas K^0 presentará también una desviación en su curva de desintegración.

El fenómeno de movimiento asimétrico tiene en cuenta así la función del pedal una corda, o pedal suave. La resonancia suele alcanzar una intensidad de unos 20 decibelios por debajo del nivel inicial del sonido inmediato, una proporción que aparentemente es agradable al oído. Esta proporción, sin embargo, no resulta adecuada para pasajes muy apacibles. Cuando en el teclado se toca "pianísimo", de forma que la amplitud del sonido inmediato se acerque a la amplitud del ruido de fondo en una sala de conciertos, la resonancia se hará inaudible. Si las notas son largas, el piano perderá su capacidad de sostener la nota y sonará como un xilofón. Para evitar esto, el piano está equipado con el pedal una corda cuya función mecánica consiste en desplazar todo el teclado de forma que un martillo sólo percuta dos cuerdas de un juego de tres. En lugar de excitar casi exclusivamente el movimiento simétrico con sólo trazas de movimiento asimétrico, el pedal una corda excita los dos tipos de movimiento casi por un igual.

¿Por qué sucede esto? La tercera cuerda, que no fue percutida por el martillo, empieza a absorber energía del puente que vibra sincrónicamente con las otras dos cuerdas. La tercera cuerda comienza de inmediato a moverse en oposición de fase con respecto a las otras dos cuerdas. De ello se sigue que hay movimiento asimétrico desde el principio. Por tanto, el nivel de resonancia se incrementa notablemente con respecto al sonido inmediato y se recobra la calidad cantabile del piano.



FUERZA RESULTANTE *(curva de color)* **sobre el puente del piano. Esa fuerza es proporcional a la suma algebraica de los desplazamientos de las cuerdas. Cuando los movimientos de las dos cuerdas** *(curvas negras)* **tienen una asimetría casi perfecta, la fuerza resultante sobre el puente está desfasada en un cuarto de ciclo con respecto al movimiento de cada cuerda, aproximadamente. El que el desplazamiento sea del orden de un cuarto de ciclo constituye una simplificación para decir que la fuerza resultante se hace mínima cuando las cuerdas alcanzan puntos de máximo desplazamiento y se hace máxima cuando las cuerdas alcanzan puntos de desplazamiento mínimo. (Ilustración de Dan Todd.)**

La resonancia que procede tanto de un movimiento asimétrico como de una polarización horizontal tiene poca intensidad comparada con el sonido inmediato. Como tal resonancia nace de irregularidades estructurales, variará a buen seguro de un modo aleatorio de nota a nota. La desafinación de las cuerdas que forman el grupo unísono es un tercer mecanismo que contribuye a la resonancia. Sin embargo, este mecanismo es ajustable; y un buen técnico en pianos probablemente variará la desafinación para compensar los efectos erráticos de las irregularidades estructurales con el fin de igualar la potencia de la resonancia de una nota a otra. Para analizar cómo afecta la desafinación a la resonancia, debe realizarse una distinción entre este fenómeno y el fenómeno de los "batidos". Si dos oscilaciones independientes cuyas frecuencias difieran ligeramente se suman, alternarán con lentitud entre un estado de refuerzo (cuando están en fase) y un estado de anulación (cuando están en oposición de fase). Para el oyente suena como una nota de frecuencia uniforme con una intensidad pulsante, que es lo que se denomina batido. Sin embargo, en un piano las dos cuerdas no vibran de una manera independiente. El movimiento del puente hace que la vibración de una cuerda afecte a la vibración de la otra. Por consiguiente, no sólo quedan afectadas notablemente las frecuencias sino también las velocidades de atenuación.

El simple movimiento de un apoyo no lleva consigo necesariamente la disipación de la energía. En ciertos sistemas físicos no se disipa energía; sólo se transfiere de un subsistema a otro. Consideremos una cuerda atada a un anillo que pueda deslizarse hacia arriba o hacia abajo, sin rozamiento, sobre una varilla fija y supongamos que el anillo está situado entre dos muelles en espiral que actúan con el fin de mantener el anillo en su posición central. Cuando la cuerda estira el anillo hacia arriba, el anillo sube; cuando la cuerda estira hacia abajo, el anillo baja. El movimiento del apoyo simula el movimiento que ejercería un cabo adicional de cuerda unida a un apoyo perfectamente rígido. Por tanto, el efecto de un apoyo "elástico" es hacer que la cuerda se mueva como si fuera más larga de lo que en realidad es y así bajar la frecuencia de la cuerda.

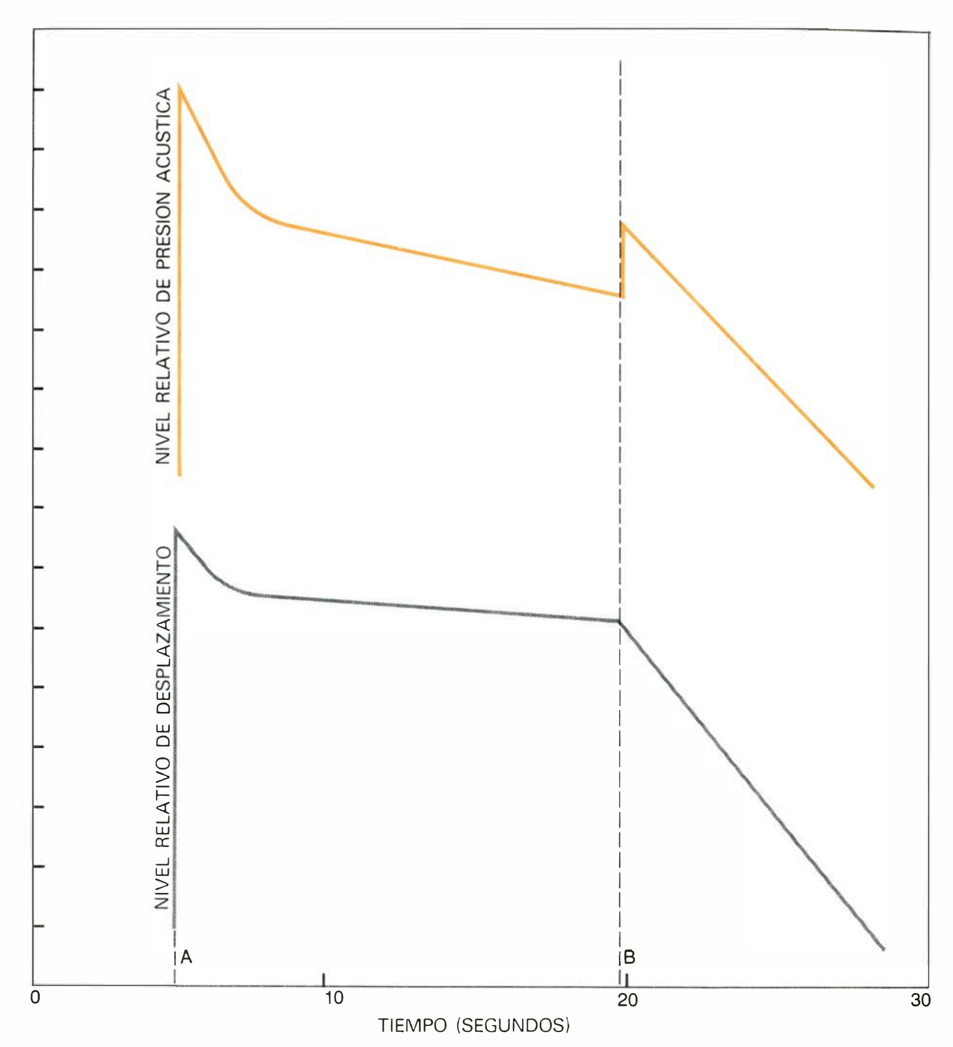
Pero un apoyo elástico no atenúa el movimiento de la cuerda, porque en el transcurso de cada ciclo completo la energía que fluye al apoyo retorna a la cuerda. A medida que la cuerda tira del anillo en su posición central, haciendo frente a la fuerza de los muelles, la cuerda trabaja sobre el anillo. Por otra

parte, conforme el anillo vuelve a su posición central, ayudado por la fuerza recuperadora de los muelles, el anillo trabaja sobre la cuerda. No existe, pues, una transferencia neta de energía.

Tampoco existiría una transferencia neta de energía en el caso de una cuerda conectada a un bloque macizo que pueda deslizarse hacia arriba y hacia abajo sin rozamiento sobre la varilla fija. En este caso, el movimiento del bloque no es gobernado por una fuerza recuperadora (ya que no hay muelles), sino por inercia. Se asume, con fines de simplificación, que la gravedad no interviene aquí. La inercia actúa manteniendo el bloque en movimiento en cualquier dirección en que se mueva. Cuando el bloque alcanza el desplazamiento máximo en un sentido, la cuerda tira hacia atrás contrarrestando la inercia al fin de reducir su velocidad e iniciar el movimiento en el sentido contrario. La inercia impulsa el bloque pasando más allá de su posición central hasta su máximo desplazamiento en este sentido. Una vez más la cuerda tira hacia atrás el bloque contrarrestando la inercia y lo mueve en el sentido inicial y el ciclo continúa repitiéndose.

El hecho que la cuerda esté tirando con frecuencia el bloque hacia atrás le hace "creer" a la cuerda que es más corta de lo que es en realidad. Por tanto, un apoyo con mucha masa incrementa la frecuencia de la cuerda. Al igual que un apoyo elástico, un apoyo de mucha masa no atenúa el movimiento de la cuerda. El trabajo que realiza la cuerda sobre el bloque mientras lo tira hacia atrás contrarrestando la inercia para cambiar su sentido es igual al trabajo que ejerce el bloque sobre la cuerda mientras estira la cuerda en el momento que la inercia impele al bloque hacia su posición central.

Los casos ideales de un anillo perfectamente elástico y un bloque totalmente macizo indican que los apoyos se pueden mover sin disipar energía. Por consiguiente, el simple movimiento del puente en un piano no es significativo de que las cuerdas pierdan energía. En realidad, el puente se parece a un tercer caso ideal: un apoyo "resistivo" donde la diferencia de fase entre el desplazamiento del apoyo y la fuerza ejercida sobre el mismo es un cuarto de ciclo. En este caso, la frecuencia de la cuerda sigue siendo la misma, pero se atenúa su movimiento. Un ejemplo de un apoyo perfectamente resistivo es un anillo cuyo movimiento no es gobernado por muelles ni por inercia, sino por rozamiento. Para vencer el rozamiento, la cuerda trabaja continuamente sobre el anillo, y así se disipa la energía de la cuerda. Que el desfase sea un cuarto de ciclo es una manera simplista de decir que, cuando el anillo llega a su desplazamiento máximo en cualquier dirección, la cuerda alcanza su desplazamiento cero (su posición central), y cuando la cuerda alcanza su desplazamiento máximo en cualquier dirección, el anillo alcanza su desplazamiento cero.



CONOCIMIENTO DEL MOVIMIENTO ASIMETRICO. Este permite construir un piano que puede introducir un acento en la parte media de una nota, que de otra forma sería sostenida. El piano se podría construir con apagadores partidos que detendrían, por separado, el movimiento de cada cuerda dentro de un grupo. En el instante *A* un martillo pone en movimiento las dos cuerdas y en el instante *B* se amortigua una de las cuerdas. La gráfica superior muestra el nivel relativo de presión acústica en función del tiempo; la inferior, el nivel de desplazamiento relativo de la cuerda no amortiguada. De *A* a *B* las curvas son típicas: primero se atenúan con rapidez y al final lentamente. Cuando una de las cuerdas se para en *B*, el fuerte movimiento asimétrico de ambas cuerdas se detiene de una manera brusca y la cuerda no amortiguada empieza de inmediato a atenuarse hasta el gradiente rápido original. La presión acústica sufre un incremento brusco porque apenas hay movimiento asimétrico para retrasar el movimiento del puente. El brusco aumento da lugar a un acento en medio de la nota.

Permítaseme explicar cómo el apoyo resistivo ideal presenta las características de la resonancia de las cuerdas desafinadas del piano. ¿Qué sucede si dos cuerdas parten con un movimiento asimétrico perfecto, pero con frecuencias no idénticas? Al principio, el soporte no se mueve, ya que el movimiento inicial de la cuerda es asimétrico. La cuerda con la frecuencia natural más aguda empieza a adelantarse en fase con respecto a la otra cuerda, y así su movimiento ya no es asimétrico puro. De ello se sigue que la cuerda ejerce una pequeña fuerza sobre el puente. La diferencia de fase entre la fuerza y el movimiento de cada cuerda es realmente un cuarto de ciclo. Las cuerdas del piano alcanzan puntos de desplazamiento máximo cuando la fuerza es mínima y puntos de desplazamiento mínimo cuando la fuerza es máxima. La premisa anterior se mantiene porque, en los puntos de desplazamiento máximo, las amplitudes de las cuerdas son opuestas y así se anulan entre sí para ejercer la fuerza mínima sobre el puente. La segunda premisa se mantiene porque en los puntos de desplazamiento mínimo las amplitudes de las cuerdas tienen el mismo signo, y así se suman para ejercer la fuerza máxima sobre el puente.

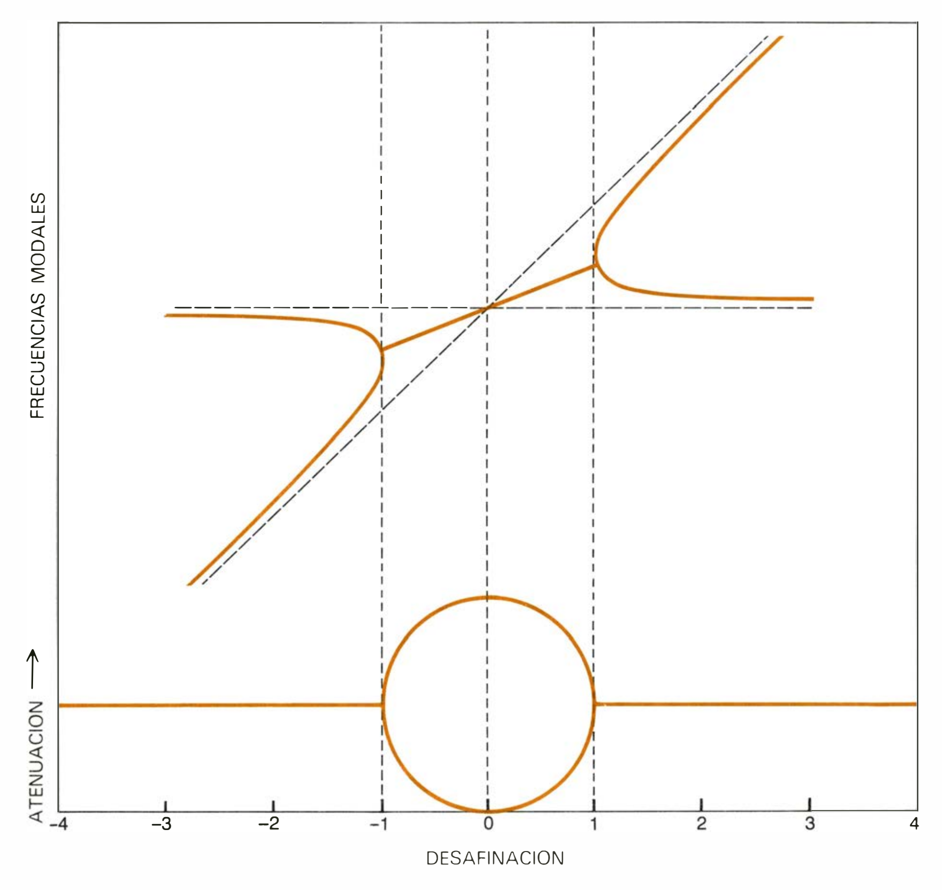
Suponiendo que el puente es un apoyo puramente resistivo, aquél desarrolla a su vez un pequeño movimiento que está desfasado en un cuarto de ciclo con respecto a la fuerza. Por tanto, el movimiento del puente está en fase con el movimiento de una de las cuerdas y está en oposición de fase con el movimiento de la otra. La cuerda que está en fase "ve" el puente como un apoyo elástico, mientras que la cuerda que está en oposición de fase "ve" el puente como un apoyo de gran masa. Esto significa que se incrementa la frecuencia de la cuerda en fase y se disminuye la frecuencia de la cuerda en oposición de fase. La cuerda con la frecuencia original más grave verá incrementada entonces su frecuencia y la otra cuerda verá su frecuencia disminuida, con lo que resultará que ambas acabarán vibrando exactamente a la misma frecuencia. Sin embargo, el gradiente de atenuación ya no es cero, como sucedía en el movimiento asimétrico puro, donde las cuerdas vibran a la misma frecuencia. Es decir, la desafinación genera un sonido de una sola frecuencia, que se atenúa poco a poco.

Si dos cuerdas parten con un movimiento perfectamente simétrico, pero con frecuencias que no son idénticas entre sí, existirá un desplazamiento de fase de una con respecto a la otra. En consecuencia, el movimiento del puente no estará exactamente desfasado en un cuarto de ciclo con respecto al movimiento de cualquiera de las cuerdas, como hubiera sucedido si las cuerdas hubieran continuado moviéndose en una forma simétrica. La frecuencia de cada cuerda se incrementa o disminuye dependiendo de si la diferencia de fase entre el movimiento de la cuerda y el movimiento del puente, está más cercano a la diferencia de fase que caracteriza un soporte elástico o más cercana a la diferencia de fase que caracteriza un soporte de gran masa. Una vez más, no se percibe el batido a medida que se aproximan las frecuencias. Como la ligera desafinación introduce un principio de movimiento asimétrico la atenuación es algo menor que en el caso perfectamente simétrico en donde las cuerdas afinadas con precisión cooperan para mover el puente.

En los casos de un movimiento inicial simétrico y un movimiento inicial asimétrico, la presencia de un acoplamiento resistivo tiende a aunar las frecuencias de las dos cuerdas, y a alterar los gradientes de atenuación. Existe un límite a la diferencia entre las frecuencias originales. A medida que crece la desafinación, la diferencia de fase entre las cuerdas aumenta, hasta que llega a un cuarto de ciclo, donde las frecuencias se separan. En el caso de mayores diferencias de frecuencia se perciben los batidos, y las velocidades de atenuación, tanto del movimiento simétrico como del asimétrico, pasan a ser iguales a los gradientes de atenuación de las cuerdas no acopladas.

Cuando dos cuerdas están afinadas su movimiento puede siempre expresarse como la superposición de modos simétricos y asimétricos. Cuando dos cuerdas están ligeramente desafinadas, su movimiento todavía se puede expresar como la superposición de dos modos: un modo casi asimétrico, cuya atenuación es pequeña, aunque no nula, y un modo casi simétrico, cuya atenuación es grande, pero no dos veces mayor que la de una sola cuerda. En ambos modos, las amplitudes de las dos cuerdas son iguales. Si el martillo percute las cuerdas al mismo tiempo y con la misma fuerza, el movimiento de perfecta simetría que excita el martillo no es un modo normal; más bien se debe considerar como una superposición, consistiendo principalmente en un modo de atenuación rápida (casi simétrica), aunque conteniendo también una parte de modo de atenuación lenta (casi asimétrica). La proporción de la mezcla depende de la diferencia de estos modos normales con respecto a la perfecta simetría y asimetría, que depende a su vez del grado de desafinación.

La principal diferencia entre la contribución a la resonancia debida a la desafinación y la contribución debida a la polarización horizontal y el movimiento asimétrico está en que un afinador de pianos experto puede ajustar la primera pero no la segunda. Creo que esto explica de forma plausible las variaciones aleatorias en la afinación al unísono que fueron observadas por Roger E. Kirk, de la Compañía D. H. Baldwin. Un afinador experto varía la desafinación de suerte que hace la resonancia uniforme y suave de nota a nota, compensando los



FRECUENCIAS DE UN PAR DE CUERDAS. Se mantienen próximas cuando los movimientos de las cuerdas están acoplados a través de un soporte puramente resistivo. La desafinación, o diferencia entre las frecuencias no acopladas, viene dada en "unidades naturales", que están relacionadas con la velocidad de atenuación de una sola cuerda. Para un par de cuerdas típico en la zona central del teclado, una unidad natural es aproximadamente un tercio de una vibración por segundo. Los trazos en el gráfico superior indican las frecuencias en ausencia de acoplamiento. El punto donde se cruzan las líneas a trazos es donde las dos cuerdas tienen la misma frecuencia. En un piano, la presencia de un apoyo resistivo hace que las frecuencias con una desafinación de +1 o −1 unidades naturales se acoplen y se mantengan a una frecuencia común. Para desafinaciones menores las frecuencias se acoplan, pero la velocidad de atenuación, que es la misma que la correspondiente a una única cuerda para desafinaciones mayores, se reparte entre las dos cuerdas *(abajo)*. (El piano prototipo se remonta a 1709.)

efectos irregulares de la polarización horizontal y el movimiento asimétrico. De este modo, el piano alcanza su belleza característica de tono que los afinadores menos expertos son incapaces de obtener. Para comprobar mi hipótesis, haría que un experto afinara el mismo piano varias veces, mientras otra persona lo fuera desafinando entre tanto. Si se manifestara siempre las mismas desafinaciones "aleatorias", éstas probarían que tales desafinaciones no eran aleatorias.

La física del piano ha llegado a un estadio en que un nuevo paso hacia adelante suscita más preguntas que soluciones. Para el investigador éste es un campo muy interesante. El método de ensayos experimentales que ha caracterizado el desarrollo de los instrumentos musicales es poco eficaz para una estructura acústica tan grande como en el caso del piano, donde la inversión necesaria para un nuevo diseño es tan elevada que frena cualquier intento de experimentación. Por esta razón, la creación de un cuadro físico detallado del funcionamiento del piano producirá un fuerte impacto en la tecnología del mismo. Incluso el actual cuadro incompleto sugiere algunas innovaciones. Por ejemplo, la comprensión del movimiento asimétrico apunta hacia la construcción de un piano que pudiera introducir un acento en medio de una nota hasta ahora sostenida.

Imagínese un piano con apagadores partidos que pudieran amortiguar, por separado, cada cuerda en un grupo unísono. Quizás un pedal especial podría controlar los apagadores partidos. Consideremos ahora un grupo unísono de dos cuerdas. Cuando se actúa la tecla correspondiente se oye una nota que tiene la mezcla típica de sonido inmediato y resonancia. Al cabo de pocos segundos, la componente simétrica del movimiento ha desaparecido y sólo se oye la resonancia. En este momento se aprieta el pedal especial que amortigua el movimiento de una de las cuerdas; en consecuencia, el fuerte movimiento asimétrico cesa y la cuerda no amortiguada empieza a atenuarse inmediatamente a la velocidad rápida inicial. De repente, se incrementa la presión sonora, pues la amplitud del movimiento del puente aumenta en ausencia del efecto retardador del movimiento asimétrico de las cuerdas. De esta manera se introduce un fuerte acento en medio de una nota normalmente sostenida. Otras formas de control de la calidad de tono se pondrán en evidencia una vez se hayan desentrañado todos los mecanismos de la física del piano.

La cabeza del cachalote

Puede representar la cuarta parte de la longitud del animal y un tercio de su peso total. En su interior, el órgano del espermaceti, lleno de aceite, puede mantener al cetáceo con flotabilidad neutra mientras bucea

Malcolm R. Clarke

Entre los grandes cetáceos el cachalote se reconoce claramente por su cabeza, que parece desproporcionadamente grande. La cabeza del cachalote puede constituir más de un tercio del peso total del animal (50 toneladas para un macho adulto medio) y más de la cuarta parte de su longitud total (18 metros de promedio). Existen buenas razones para esta desproporción aparente. El cráneo del cachalote supone alrededor del 12 por cien del peso de la cabeza; el restante 88 por cien está formado principalmente por una peculiar característica anatómica, situada en el hocico del cetáceo, sobre la mandíbula superior: el órgano del espermaceti. Este órgano es una complicada masa de músculo y de tejido conjuntivo lleno de aceite. El aceite es el origen del nombre anglosajón *(sperm whale)* del cachalote; en un macho grande, el órgano puede contener hasta cuatro toneladas de aceite de espermaceti.

Es evidente que un órgano tan enorme debe desempeñar alguna función importante en la vida del cachalote. Durante mucho tiempo se ha especulado acerca de la utilidad de la gran "caja", como la suelen denominar los balleneros. Incluso la misma estructura del órgano del espermaceti era poco conocida hasta la última década. Anteriormente, las suposiciones sobre su función podían basarse únicamente en generalidades y en lo que se sabía de los órganos similares, pero mucho más pequeños, de algunos otros cetáceos odontocetos. En la actualidad se sabe lo suficiente acerca de este órgano para que sea posible sugerir su función principal: permitir que el cachalote conserve una flotabilidad neutra cuando está sumergido.

¿Por qué no se conocieron mucho antes los detalles anatómicos del órgano del espermaceti? Parte de la respuesta es que los pioneros en este campo obtenían sus datos disecando fetos de cachalote. Pero resulta que tanto el cráneo como el hocico del feto tienen proporciones muy distintas de las de los mismos componentes en la cabeza del cetáceo adulto. Otra parte de la respuesta es que no es tarea fácil disecar un cachalote adulto. Sin los medios y las herramientas de la industria ballenera, como plataformas de despiece, montacargas hidráulicos, sierras hidráulicas de cinco metros y cuchillas de despiece afiladas como navajas, la disección sería imposible. Incluso con estos medios y la cooperación de los balleneros profesionales no resulta una tarea fácil.

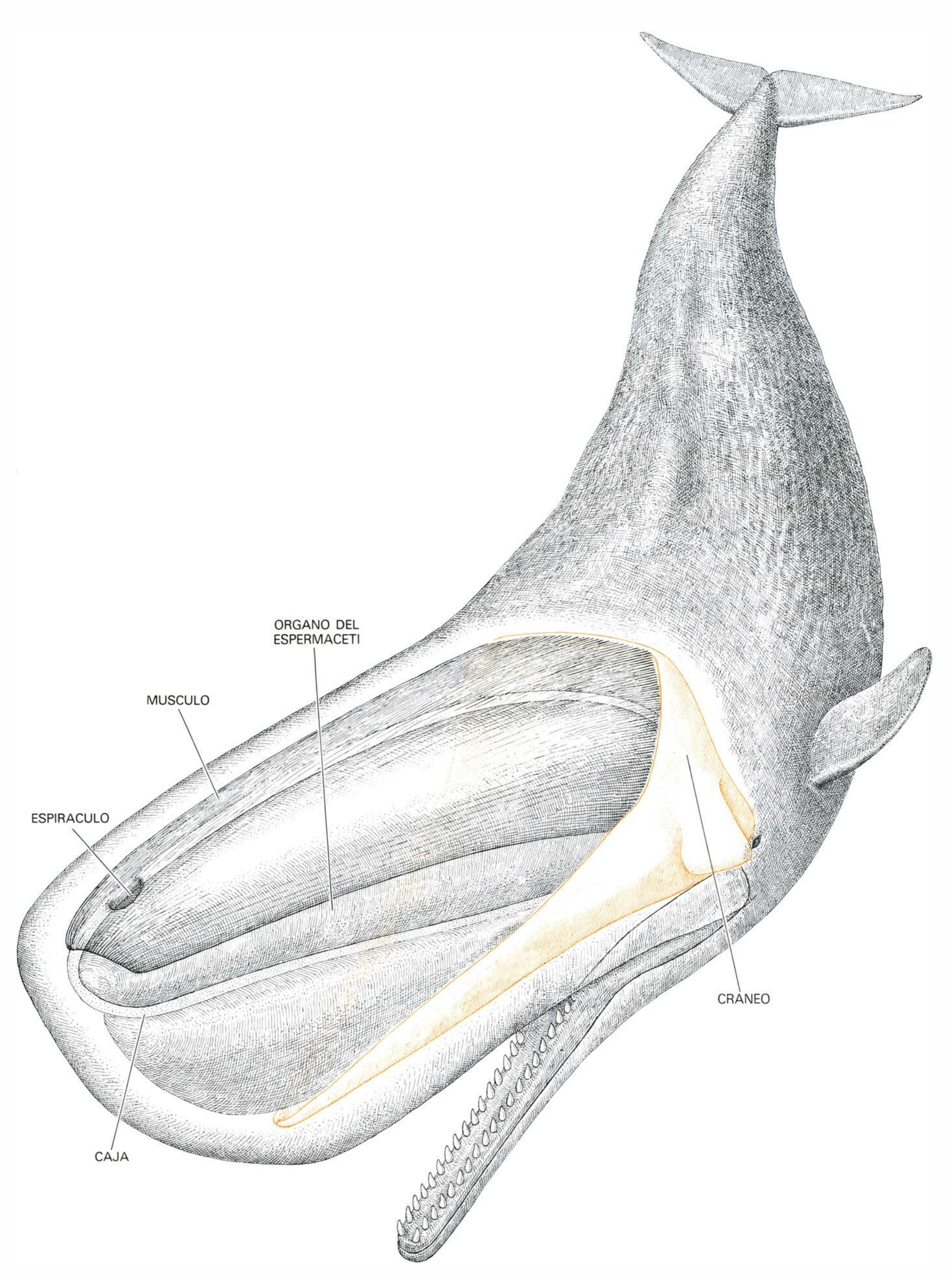
En la plataforma de despiece se cortan, se hacen girar y se vuelven del revés una y otra vez masas de media tonelada de grasa, carne y fibra, hasta que el sentido de orientación del observador se pierde fácilmente. No es raro que en esta misma década se haya publicado un libro que presenta invertidas las principales estructuras de la cabeza del cachalote. Sólo después de observar y fotografiar muchos de los cachalotes que se cortan en la plataforma de despiece, puede esperarse clarificar la anatomía del hocico del cachalote adulto, y ésta es la tarea que me propuse. Mi trabajo se vio facilitado en gran manera por la captura de un pequeño adulto cuya cabeza cortaron para mí los balleneros en una serie de secciones transversales de 20 centímetros de grosor. Pude entonces fotografiar las secciones y medirlas en detalle.

Se han atribuido muchas funciones al órgano del espermaceti. Unos dicen que se trata de un sistema para la producción y la concentración de sonidos (y para recibirlos); otros, que es un sistema para mover el aire entre los pulmones del cetáceo y sus orificios nasales cuando éste se halla sumergido a cierta profundidad; otros, que no es más que un sistema para abrir y cerrar los largos conductos nasales del cachalote; y otros, que se trata de un sistema para absorber nitrógeno de la corriente sanguínea durante las inmersiones profundas; incluso ha habido quien lo propuso como un sistema de ataque y defensa.

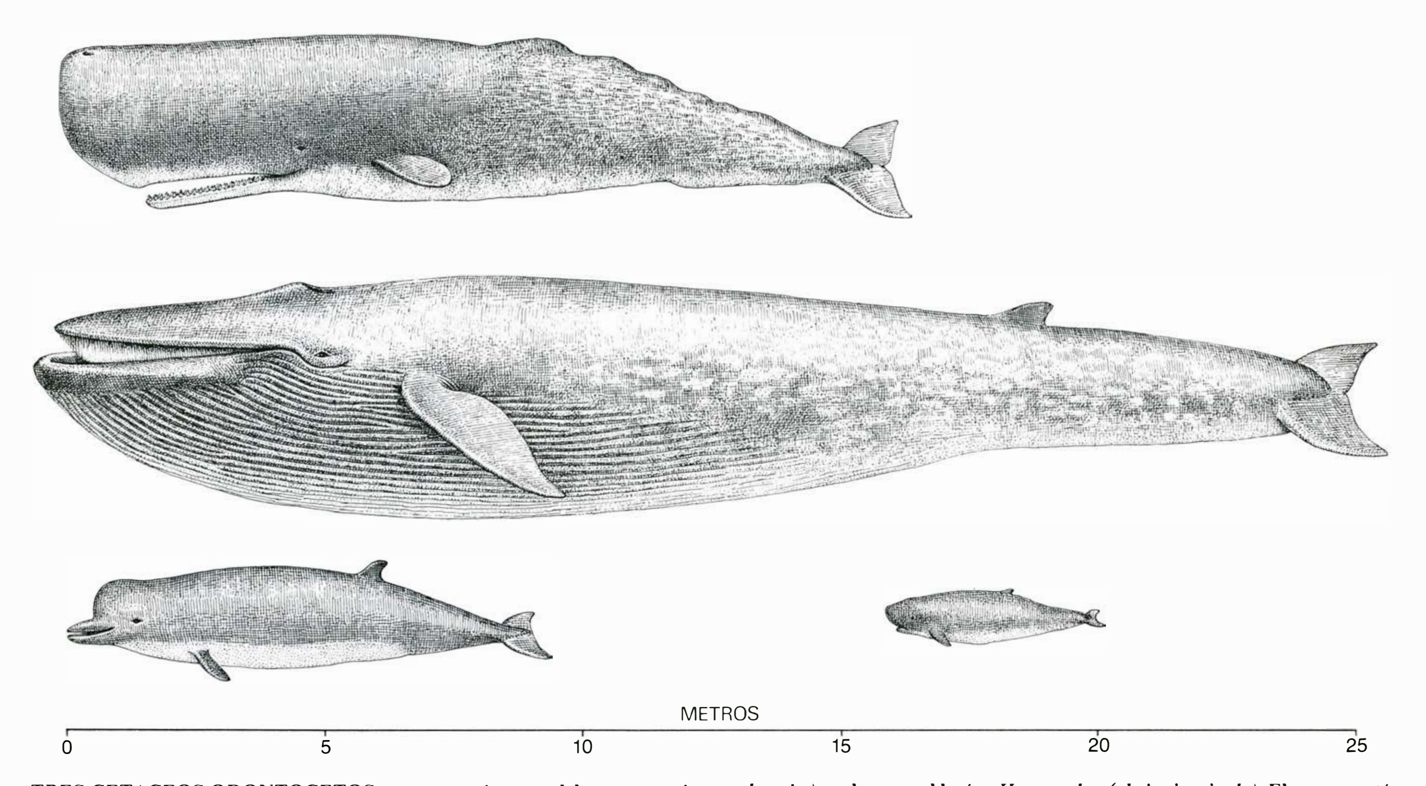
Ciertamente, este órgano complejo puede tener más de una función. Las variaciones en la composición química del aceite de espermaceti en distintas partes del órgano sugieren que una de estas funciones bien pudiera ser la conducción (o la concentración o el enfoque) del sonido generado por el cetáceo. Al mismo tiempo, es difícil aceptar la hipótesis de un investigador que opina que este enfoque puede concentrar el sonido de manera lo suficientemente intensa como para aturdir a los calamares, que constituyen la presa principal del cachalote. Algunas de las otras posibilidades, aunque parecen razonables, son difíciles de demostrar.

Si bien el conocimiento de la estructura y de las proporciones de la cabeza del cachalote adulto es un preliminar necesario para el estudio de la función del órgano del espermaceti, también debe saberse algo sobre la biología y el comportamiento de los cachalotes, sobre todo las diferencias que existen entre estos cetáceos odontocetos y los demás que carecen de un gran hocico. Además, la gran cantidad de aceite que contiene el órgano del espermaceti sugiere que el propio aceite debe servir para alguna función especial; hay que conocer algo de las propiedades físicas del aceite para comprender el papel que desempeña en la vida del cachalote.

Empecemos nuesta búsqueda de pistas sobre la función del hocico del cachalote revisando la biología y el comportamiento del animal. Un hecho que resulta evidente es que los cachalotes son insólitos, aunque no únicos, en el conjunto de los cetáceos odontocetos al hallarse distribuidos por todo el mundo. Durante siglos se les ha cazado en todos los mares, desde los 60 grados de latitud norte hasta los 40 grados de latitud sur. A raíz del desarrollo de las modernas técnicas balleneras, el cachalote ha sido perseguido asimismo hasta las elevadas latitudes del Antártico. Junto a su primo misticeto, la ballena franca, el cachalote era de gran importancia cuando la in-



HOCICO DEL CACHALOTE, que puede pesar más de 12 toneladas. El hocico está ocupado en gran parte por el órgano del espermaceti, una complicada masa oblonga de tejido conjuntivo embebido en aceite que se halla rodeada por capas de músculo y de grasa y que está encerrada por una membrana llamada "caja". El órgano se halla encajado en una larga depresión cónica formada a partir de las mandíbulas superiores *(color claro).*



TRES CETACEOS ODONTOCETOS que poseen órganos del espermaceti se comparan en esta ilustración con el mayor de todos los cetáceos. Se trata del cachalote, *Physeter catodon (arriba),* el cachalote enano, *Kogia (abajo, derecha),* y el gran calderón, *Hyperoodon (abajo, izquierda).* El mayor cetáceo es el rorcual azul, *Balaenoptera musculus,* del suborden Misticetos, es decir, de los cetáceos con barbas (el de Odontocetos porta dientes).

dustria ballenera dependía del arponeo a mano desde barcas de remos: a diferencia de muchos otros cetáceos, estos dos permanecían a flote después de haber sido muertos.

El alimento del cachalote está compuesto casi por entero por los "Concorde" del grupo de los moluscos, los calamares. Para capturar a estos veloces moradores de las profundidades, el cachalote realiza inmersiones profundas y permanece largos períodos sumergido. Un cachalote grande se sumerge típicamente durante 50 minutos de un ciclo de buceo que dura una hora. A lo largo del intervalo de 10 minutos de estancia en superficie entre inmersiones, el cetáceo realiza de 50 a 60 inspiraciones de aire. Se han observado inmersiones de más de 50 minutos; la inmersión récord es de algo más de 80 minutos.

Los cachalotes no sólo permanecen sumergidos durante largos períodos, sino que con frecuencia llegan más allá de los 1000 metros de profundidad. Mediante el sonar se observó a un cachalote que descendía a más de 2250 metros; y pueden deducirse descensos aún más profundos a partir de la presencia de tiburones bentónicos en el estómago de un cachalote capturado en una zona en que la profundidad del fondo superaba los 3000 metros.

¿Por qué razón necesita bucear tan profundamente un cachalote? En todas las profundidades de todos los océanos se encuentran calamares. Constituyen el principal alimento de muchos otros vertebrados de respiración aérea: aves marinas, mamíferos marinos como las focas, y cetáceos odontocetos menores como las marsopas y los delfines. Quizá la razón del comportamiento buceador del cachalote sea que cuanto más profundamente pueda descender un comedor de calamares más atrás dejará a sus competidores y mayor será la cantidad de calamares a su disposición. Ciertamente, el cachalote puede capturar calamares abisales que están fuera del alcance de las aves marinas, de las focas o de los delfines. El cachalote puede asimismo alcanzar el fondo en la zona en que el suelo marino desciende abruptamente hacia la llanura abisal desde el borde de la plataforma continental. Aquí, en el talud continental, a profundidades que van desde los 200 a los 3000 metros, muchos calamares depositan su puesta y pueden encontrarse en gran número y concentraciones densas. Pocos animales de respiración aérea que no sean el cachalote pueden confiar en alcanzar una pitanza tan fácil.

Una peculiaridad del ciclo de buceo del cachalote es que con frecuencia el animal vuelve a la superficie en un punto situado a unos pocos cientos de metros de aquél en el que empezó su inmersión. La razón no estriba en que su tiempo de inmersión haya sido ocupado por un descenso y un ascenso lentos. Se sabe que los cachalotes descienden a una velocidad de unos cuatro nudos (120 metros por minuto) y ascienden a una velocidad de unos cinco nudos, de modo que un viaje de ida y vuelta a una profundidad de 1000 metros no duraría más de 15 minutos. La duración de una inmersión profunda es unas tres veces superior. Tanto la duración de la inmersión como el hecho de que el lugar de emersión del cachalote se halle cerca del lugar de inmersión sugieren que cuando el cachalote llega al fondo de su zambullida muchas veces permanece casi inmóvil en el agua.

Muchos calamares, aunque no todos, son nadadores rápidos en distancias cortas. Por ello cabría preguntarse por qué razón un cachalote permanece quieto en las profundidades en lugar de perseguir activamente a sus veloces presas. La mandíbula inferior del cachalote es larga y estrecha. Es la mandíbula de un mordedor; incluso estando abierta presenta muy poca resistencia al agua. Quizá la estrategia de caza del cachalote se basa menos en la persecución activa y más en la lenta suspensión entre dos aguas, seguida de una rápida embestida al banco de calamares que pase cerca. A estas profundidades de caza apenas penetra la luz solar, pero la mayoría de calamares sobre los que el cachalote hace presa son luminiscentes. En sus esfuerzos para capturar estos veloces invertebrados, un cetáceo inmóvil y silencioso bien pudiera tener más ventajas que un cetáceo nadador.

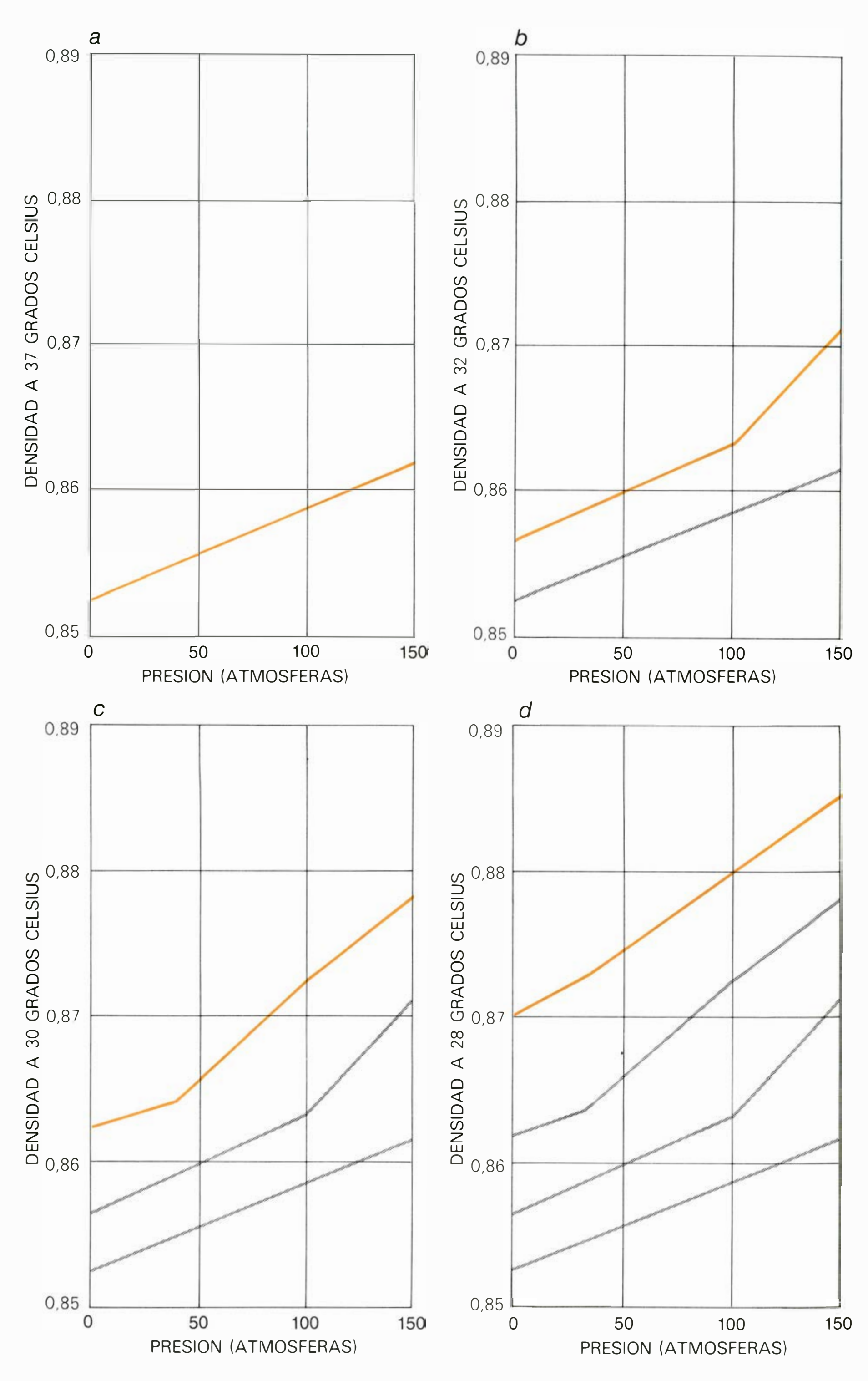
El cetáceo sólo puede permanecer inmóvil en el agua si está muy cerca de tener una flotabilidad neutra, es decir, si tiene la misma densidad que el agua que le rodea. En los tipos más antiguos de submarinos fabricados por el hombre, que desplazaban de 1600 a 2000 toneladas, la flotabilidad tenía que controlarse dentro de un margen de 40 litros de agua (entre dos y tres cienmilésimas del peso de desplazamiento) para conseguir que la nave permaneciera inmóvil en el agua con fines de escucha. Muchos animales que viven en el agua pueden alcanzar asimismo la flotabilidad dentro de límites muy finos, lo cual les permite permanecer inmóviles en las profundidades en que viven. Por ejemplo, algunos peces contrarrestan el efecto de hundimiento de aquellos tejidos corporales que son más densos que el agua almacenando grasas de baja densidad; otros peces consiguen la flotabilidad neutra mediante una vejiga natatoria llena de gas. Muchos calamares hacen lo mismo al sustituir los densos iones sodio de su cuerpo por iones amonio, menos densos.

Así pues, en sus aspectos biológico y de comportamiento, el cachalote presenta algunos rasgos insólitos. Entre ellos está el comportamiento migrador que lleva a los machos de las aguas ecuatoriales a las polares, las inmersiones profundas y prolongadas, la capacidad de permanecer inmóvil mientras está sumergido y la propiedad de flotar cuando está muerto. Ninguna de estas características es exclusiva del cachalote, pero sólo el cachalote parece combinarlas todas a la vez.

De estas cuatro características, dos corresponden a la flotabilidad, y las otras dos (extensa distribución, tanto horizontal como verticalmente) implican cambios en las condiciones del agua circundante que están acompañados por cambios en la flotabilidad. Aunque el cachalote podría llevar la cantidad adecuada de grasa o de aire para tener una flotabilidad neutra en una localidad geográfica y a una profundidad determinadas, si el cetáceo dependiera de un sistema tan estático para el control de la flotabilidad, las distintas densidades del agua a otras profundidades y en otras localidades geográficas lo empujarían hacia abajo o hacia arriba.

¿Es ésta la pista que estábamos buscando? ¿Es el órgano del espermaceti un mecanismo para el control de la flotabilidad para un amplio rango de condiciones? Si es así, el órgano debe ser capaz de variar su densidad. ¿Cómo puede hacerlo? Sólo existe una sustancia en el órgano en grandes cantidades que sea asimismo capaz de sufrir un cambio de densidad sustancial: el propio aceite de espermaceti. En realidad, desde hace tiempo se sabe que este aceite tiene propiedades distintas de los demás aceites de cetáceos. Cuando se extrae aceite líquido de la cabeza de un cachalote muerto y se expone a la temperatura del aire ambiente, pronto pierde su transparencia y se transforma en un sólido cristalino blando.

Mediciones de temperatura en cachalotes recién muertos indican que cuando los animales descansan en superficie la temperatura del aceite de espermaceti es de 33 grados Celsius. El aceite empieza a cristalizar, es decir, a congelarse solidificándose, cuando su temperatura des-



EL ACEITE DE ESPERMACETI varía de densidad según la temperatura. Estos gráficos indican la densidad del aceite a temperaturas progresivamente menores: a 37 grados Celsius *(a)*, a 32 grados *(b, con la lectura a 37 grados, que se incluye para su comparación)*, a 30 grados *(c)* y a 28 grados *(d)*. La presión también afecta a la densidad del aceite. Cada 10 metros de profundidad adicional añade una atmósfera (1,033 kg) de presión; así, a una profundidad de 500 metros la presión es de 50 atmósferas. Las pendientes de los cuatro gráficos siguen la densidad creciente del aceite hasta el equivalente a 1500 metros. Los valores proceden de un estudio realizado en el British National Physical Laboratory.

ciende por debajo de los 31 grados C. A diferencia de la cristalización del agua, que es casi instantánea en el punto de congelación, la cristalización del aceite de espermaceti es un proceso gradual que no termina hasta que la temperatura desciende varios grados. Cuando el aceite de espermaceti se congela, se hace más denso y ocupa por tanto menos volumen. Y al ocupar un volumen menor, desplaza menos cantidad de agua circundante y flota menos.

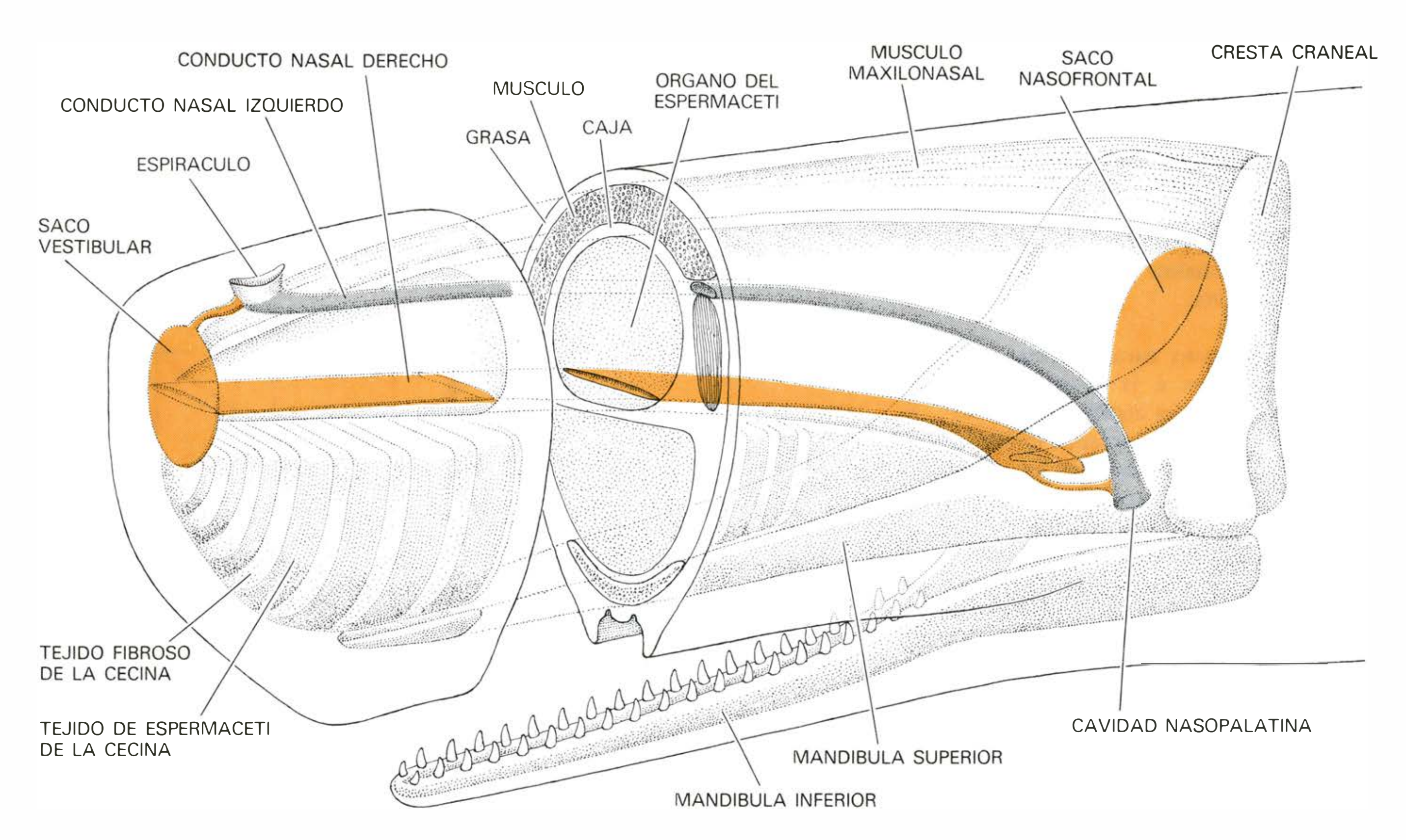
Entonces, si pudiera hacerse variar la temperatura del aceite de espermaceti, los cambios de densidad que acompañan a los cambios de temperatura serían suficientes para permitir que el cachalote controlara su propia flotabilidad. ¿Son fisiológicamente posibles estos cambios de temperatura? Desde luego, el problema que se presenta es de pérdida y ganancia de calor. Teniendo en cuenta todo esto, examinemos ahora la anatomía del órgano del espermaceti.

Los tejidos que albergan el aceite de espermaceti en el hocico del cachalote tienen una densa red de capilares irrigados con sangre por grandes arterias que penetran en el hocico por detrás. La circulación de la sangre arterial es, por lo tanto, el principal sistema de aporte de calor al aceite. La propia circulación a nivel capilar constituye también el principal distribuidor de calor en el interior de cada bloque de tejido de espermaceti; cuando el tejido se enfría localmente, el movimiento de la sangre a través de los capilares colabora a extender el efecto refrigerante.

Además, las arterias y venas mayores del hocico del cachalote se sitúan unas junto a las otras; este sistema de contracorriente coadyuva al enfriamiento del tejido de espermaceti mediante el intercambio de calor entre la sangre arterial que llega, más cálida, y la sangre venosa que se va, más fría. Las arterias que suministran la sangre al hocico están rodeadas, en particular en el punto por donde pasan a través del cráneo, por una densa red de venas que se llevan la sangre más fría. El intercambio de calor puede mantener una notable diferencia de temperatura entre el hocico (donde por lo general la sangre está a menos de 34 grados Celsius) y el resto del cuerpo del cetáceo (donde normalmente la sangre se halla por encima de los 37 grados Celsius).

El cachalote puede perder calor a través de los tejidos superficiales del hocico mediante la conducción "pasiva" a través de la capa de grasa y de la piel, o bien mediante el transporte "activo" de calor: la circulación de la sangre hasta las papilas, diminutas estructuras digitiformes del interior de la piel. Como veremos, la conducción pasiva por sí sola es un proceso demasiado lento para la consecución de la flotabilidad neutra en el tiempo disponible. El transporte activo de calor es una posibilidad más prometedora.

La anatomía de la cabeza del cachalote sugiere todavía otro mecanismo de pérdida de calor: los conductos nasales de este cetáceo. Los conductos nasales asimétricos son característicos de los Odontocetos, pero ningún otro odontoceto tiene conductos nasales parecidos a los del cachalote. El más simple de estos conductos, el izquierdo, corre hacia atrás desde la cavidad situada bajo el espiráculo único del animal, se curva para pasar por el lado izquierdo del órgano del espermaceti y penetra en el cráneo



CONDUCTOS NASALES del cachalote. No sólo son asimétricos sino que se hallan íntimamente asociados con el órgano del espermaceti. El conducto nasal izquierdo *(gris obscuro)* es el más sencillo de los dos. Empieza en una cavidad situada bajo el espiráculo, se curva para pasar a lo largo del lado izquierdo de la caja del espermaceti y termina en la cavidad nasopalatina del cráneo. El conducto nasal derecho *(color)* se inicia asimismo bajo el espiráculo; después corre hacia adelante y se ensancha para formar un saco orientado verticalmente, el saco vestibular, en el extremo anterior de la caja del espermaceti. Una estrecha abertura horizontal en la parte posterior del saco permite la continuación del conducto hacia atrás a través del interior del órgano del espermaceti, hasta que el conducto se acerca a la cresta en forma de cuchara del cráneo del cetáceo. Aquí, el conducto nasal da origen a un segundo saco orientado verticalmente, el saco nasofrontal, situado en la parte posterior de la caja del espermaceti, antes de estrecharse para entrar en la cavidad craneal que comparte con el conducto nasal izquierdo. La ruta que sigue el conducto nasal derecho tiene unos cinco metros de longitud, y alcanza más de un metro de ancho. A lo largo de la mayor parte de su longitud se extienden hacia la pared del conducto nasal vasos procedentes de la red de capilares del tejido del aceite de espermaceti. El músculo maxilonasal va desde la cresta craneal hasta la mitad anterior de la caia

exactamente delante de la cavidad craneal. El cachalote respira a través de este tubo muscular, que puede expandirse hasta que su sección transversal se hace circular.

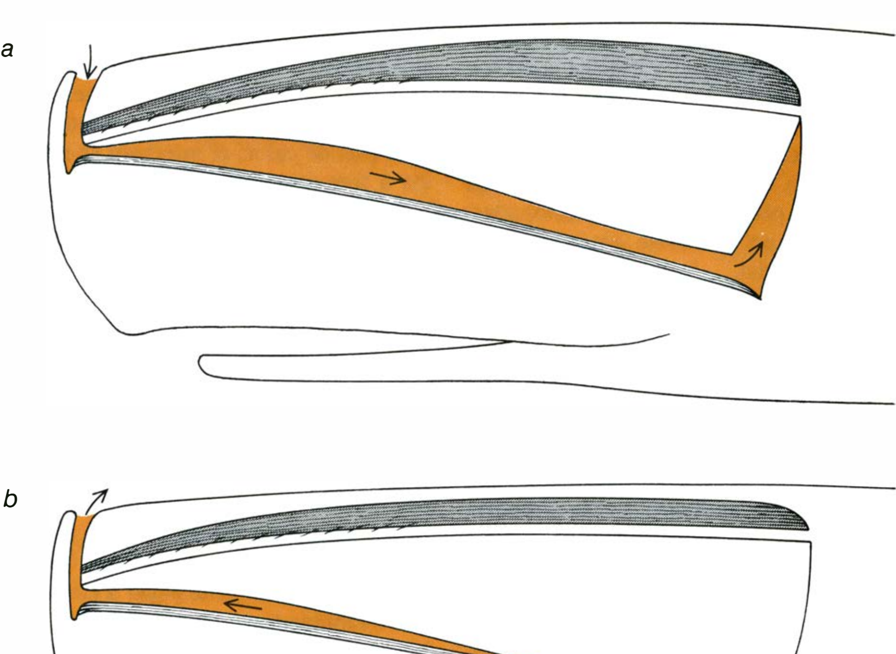
El conducto nasal derecho tiene una morfología totalmente distinta. Aunque empieza también en la cavidad situada debajo del espiráculo, corre primero hacia adelante y es un conducto tubular de pequeño diámetro. Después se ensancha y se aplasta mientras pasa hacia abajo para formar una cámara amplia y plana, el saco vestibular, situado en la parte frontal del hocico, exactamente debajo de la capa de grasa. Una amplia abertura horizontal en la pared posterior del saco, llamada a menudo "boca de mono", da paso a la continuación del conducto nasal, tubo ancho y plano que corre hacia atrás a lo largo de todo el hocico hasta un punto situado frente al cráneo.

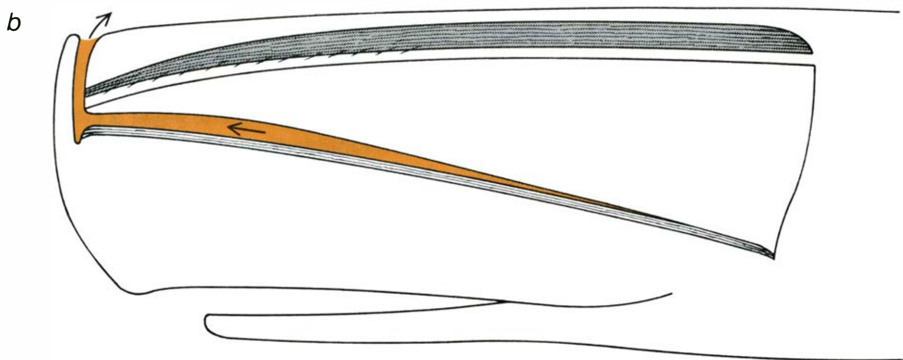
Aquí el amplio paso se estrecha al entrar en el cráneo y unirse al conducto nasal izquierdo en una cavidad común. Exactamente delante de este estrechamiento, el conducto nasal derecho se abre hacia arriba para conectar con un segundo saco, el nasofrontal, situado por encima de la parte central de la cresta craneal.

El conducto nasal derecho suele tener, pues, más de un metro de anchura; su trayecto tortuoso a lo largo del hocico mide cinco o más metros de longitud. No sólo pasa a través del centro del órgano del espermaceti, sino que además sus dos sacos cubren los extremos anterior y posterior del órgano.

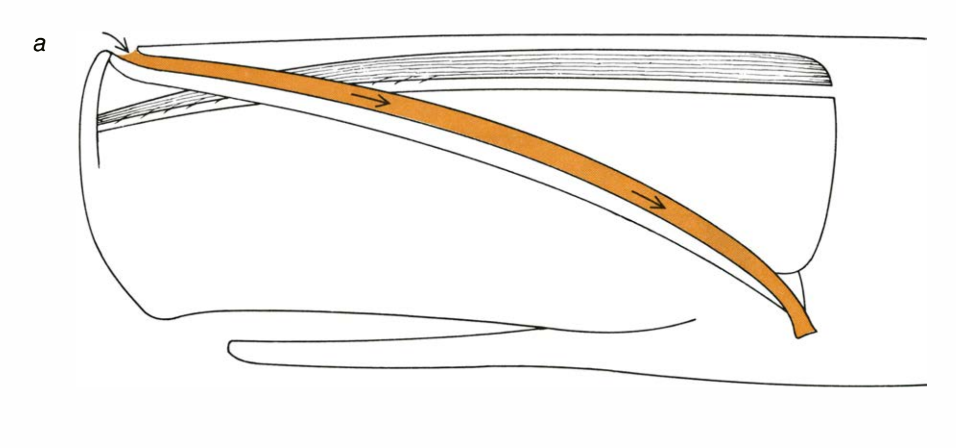
La parte interior del conducto nasal derecho está tapizada con una delgada capa de tejido negro. Debajo de la capa negra se encuentran otras dos capas de tejido blanco, primero una capa elástica y después una fibrosa; las tres capas juntas forman una pared que tiene de 0,6 a 1 milímetro de espesor. El tejido del espermaceti se halla directamente en contacto con la pared del conducto nasal por todas partes, y los capilares procedentes del tejido del espermaceti entran en la capa blanca elástica.

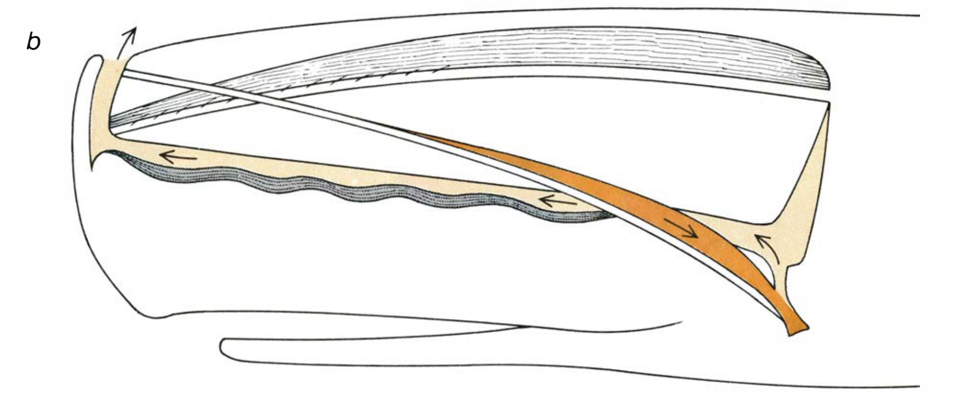
La relación íntima entre el conducto nasal derecho del cachalote y el órgano del espermaceti es tal que, si por el conducto nasal penetra agua de mar, el aceite de espermaceti se enfriará considerablemente. Puede calcularse la tasa de enfriamiento del aceite para cualquier temperatura determinada del agua de mar (sobre la base del intercambio de calor entre la sangre de los capilares y el agua de mar, más fría) a partir de la superficie total y del espesor de la pared del conducto nasal. Se puede calcular, además, la tasa de pérdida de calor a





ENTRADA DE AGUA, a través del conducto nasal derecho *(a)*, mediante la contracción del músculo principal del hocico *(gris obscuro)*. La acción muscular ensancharía el conducto *(color)* para que pudiera entrar el agua; el proceso sería facilitado por músculos del interior del tejido del espermaceti. La relajación de éstos permitiría que el conducto se estrechara de nuevo *(b)*, expulsando así el agua.





RUTA ALTERNATIVA DE ENTRADA, constituida por el conducto nasal izquierdo *(color)*. La contracción de un músculo menor, el músculo del obturador nasal, ensancharía el conducto izquierdo y haría entrar agua en la cavidad nasopalatina *(a)*. Posteriormente *(b)*, la acción de las fibras musculares en el suelo del conducto nasal derecho *(gris)* podría bombear el agua hacia adelante a través del órgano del espermaceti, y hacia afuera a través del espiráculo. Esta vía exigió la disección del animal.

través de la piel del hocico del cetáceo, a través del conducto nasal derecho o a través de las dos áreas de intercambio de calor combinadas, y determinar así el tiempo necesario para que el cachalote alcance la flotabilidad neutra en una gama de profundidades seleccionadas, ya sea en un ambiente antártico o en otro ecuatorial.

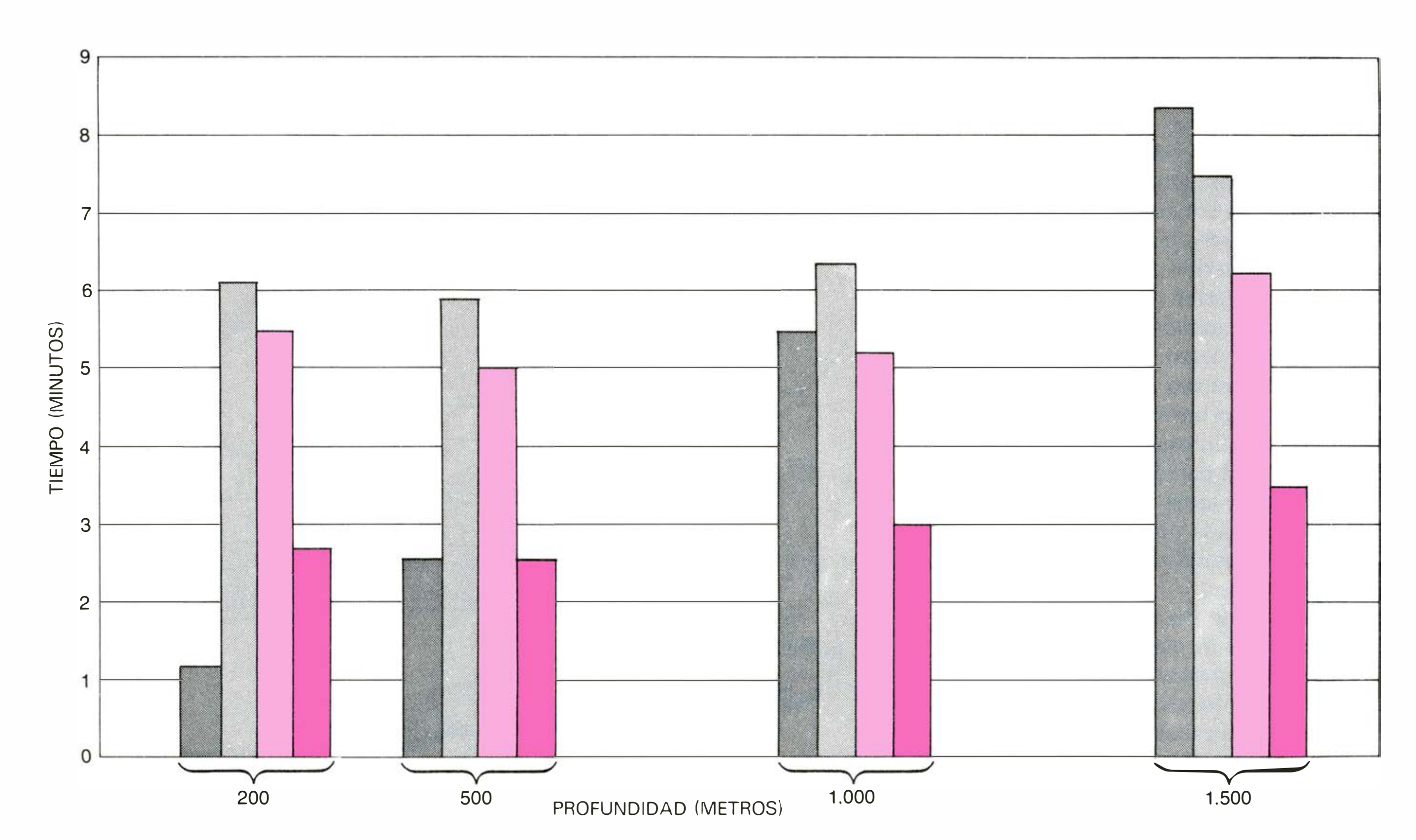
En cualquier ambiente, con valores mínimos en el cálculo, el conducto nasal derecho resulta ser un intercambiador de calor más lento que la piel del hocico del cachalote. Sin embargo, la diferencia calculada entre las dos superficies probablemente no es significativa; el conducto nasal es elástico y su pared podría expandirse para presentar una mayor superficie de intercambio de calor que la mínima que yo he estimado.

Mis cálculos demuestran que un cachalote de 30 toneladas que explote simultáneamente las dos zonas intercambiadoras de calor puede ajustarse a la flotabilidad neutra en menos tiempo del que normalmente requiere para descender nadando a una profundidad de 500 metros. En inmersiones de 200 a 1000 metros de profundidad, e intercambiando calor exclusivamente a través de la piel del hocico, el cetáceo alcanzaría la flotabilidad neutra a los cinco minutos de llegar a la profundidad deseada. Si actuaran a la vez las dos zonas intercambiadoras de calor, el intervalo se acortaría a tres minutos.

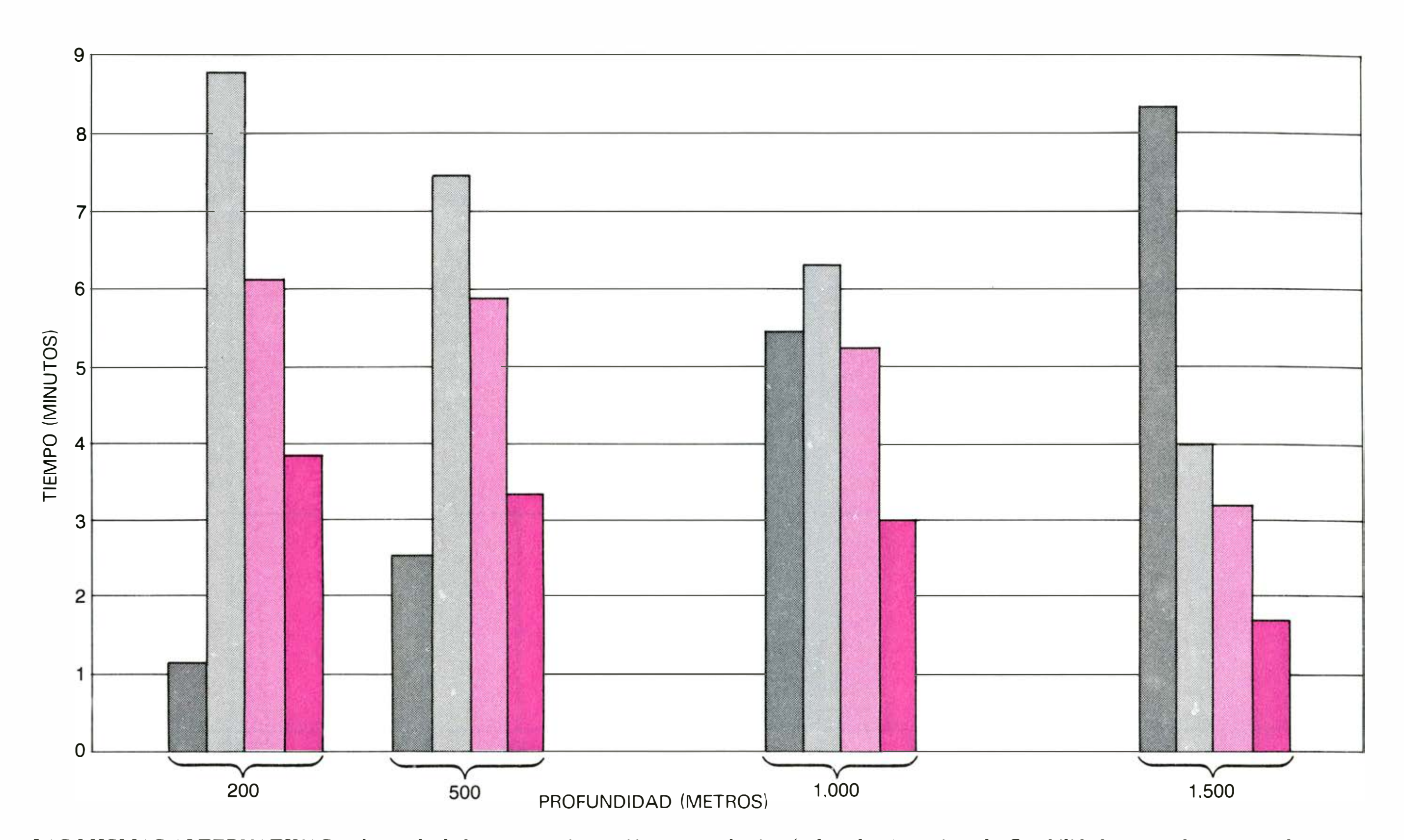
En el Antártico, y debido a que las temperaturas inferiores del agua implican una mayor flotabilidad en la superficie y cerca de ella, el tiempo calculado para alcanzar la flotabilidad neutra a una profundidad de 100 metros, actuando cualquiera de los dos intercambiadores de calor por separado, llega a ser de alrededor del 20 por ciento del tiempo total de inmersión, incluso en el caso de un buceo prolongado. A primera vista, esto puede parecer una inversión de tiempo demasiado grande para que valga la pena llegar a conseguir la flotabilidad neutra. Existe, sin embargo, un factor de compensación: al nadar en las frías aguas superficiales de las latitudes antárticas, el cachalote puede mantener la temperatura de su aceite de espermaceti a un nivel inferior a los 33 grados C característicos de las aguas ecuatoriales. Si lo hace así, el enfriamiento hasta la flotabilidad neutra sería más rápido. Además, cuando el odontoceto se halla sumergido durante el buceo, si utiliza repetidamente el intercambiador de calor nasal puede acelerar la consecución de la flotabilidad neutra. Mis cálculos indican que, por debajo de los 200 metros, llenar el conducto nasal con agua de mar no una vez, sino dos, sería más que suficiente para intercambiar la cantidad de calor necesaria.

¿De qué manera se introduce el agua de mar en el conducto nasal derecho del cachalote? Rodeando la pared externa que encierra los tejidos del espermaceti existen grandes músculos. Corren desde la mitad anterior de la "caja" hasta inserciones sobre la cresta del cráneo. Su contracción bastaría para levantar la parte anterior de la caja, alzando así la mitad superior del conducto nasal. La misma contracción abriría asimismo el extremo anterior del conducto nasal y succionaría agua de la cavidad situada bajo el espiráculo. La relajación de los músculos expulsaría el agua.

¿A qué distancia puede desplazarse el agua de mar a lo largo del conducto nasal derecho? Existen fibras musculares adicionales en el suelo del conducto, y otras más que corren hacia delante en el



TRES ALTERNATIVAS para enfriar el aceite de espermaceti cuando un cachalote bucea en el ecuador. Como puede apreciarse, tardan períodos distintos de tiempo. Los cálculos corresponden a un cachalote de 30 toneladas que realiza la inmersión con los pulmones llenos. La primera columna *(gris)* en cada uno de los cuatro grupos indica el tiempo que el animal necesita para nadar hasta la profundidad indicada a una velocidad de cinco nudos. Cuando la inmersión es a 200 metros, incluso el intercambio de calor máximo a través del conducto nasal derecho y de la piel del hocico a la vez *(color)* debe continuar durante más de un minuto adicional antes de que se alcance la flotabilidad neutra. El intercambio de calor a través del conducto nasal derecho *(gris claro)* o a través de la piel del hocico *(color claro)* tendría que continuar durante más tiempo todavía. Con inmersiones más profundas, la tendencia favorece la consecución de la flotabilidad neutra durante el tiempo en que el cachalote está nadando hacia el fondo. En una inmersión a 500 metros, el tiempo necesario para conseguir la flotabilidad neutra mediante el máximo intercambio de calor es igual al tiempo de descenso. En una inmersión a 1500 metros, hasta la forma menos eficaz de intercambio de calor consigue la flotabilidad neutra durante el descenso.



LAS MISMAS ALTERNATIVAS se han calculado para una inmersión en aguas polares; también aquí el cachalote pesa 30 toneladas y nada hacia abajo con los pulmones llenos y a una velocidad de cinco nudos. El intercambio máximo de calor *(color)* consigue casi la flotabilidad neutra durante un descenso a 500 metros. El intercambio de calor a través de la piel del hocico *(color claro)* consigue la flotabilidad neutra durante un descenso a 1000 metros. Incluso la forma menos eficiente de intercambio de calor *(gris claro)* consigue la flotabilidad neutra cuatro minutos antes de que alcance los 1500 metros. (Al desplazarse entre latitudes polares y ecuatoriales, experimenta un cambio térmico en las aguas de hasta 26 grados Celsius.)

interior del tejido del espermaceti a partir de la pared frontal del saco nasofrontal. La contracción de estas fibras mantendría deprimida la mitad inferior del conducto nasal derecho, asegurando el movimiento del agua por lo menos hasta el saco nasofrontal.

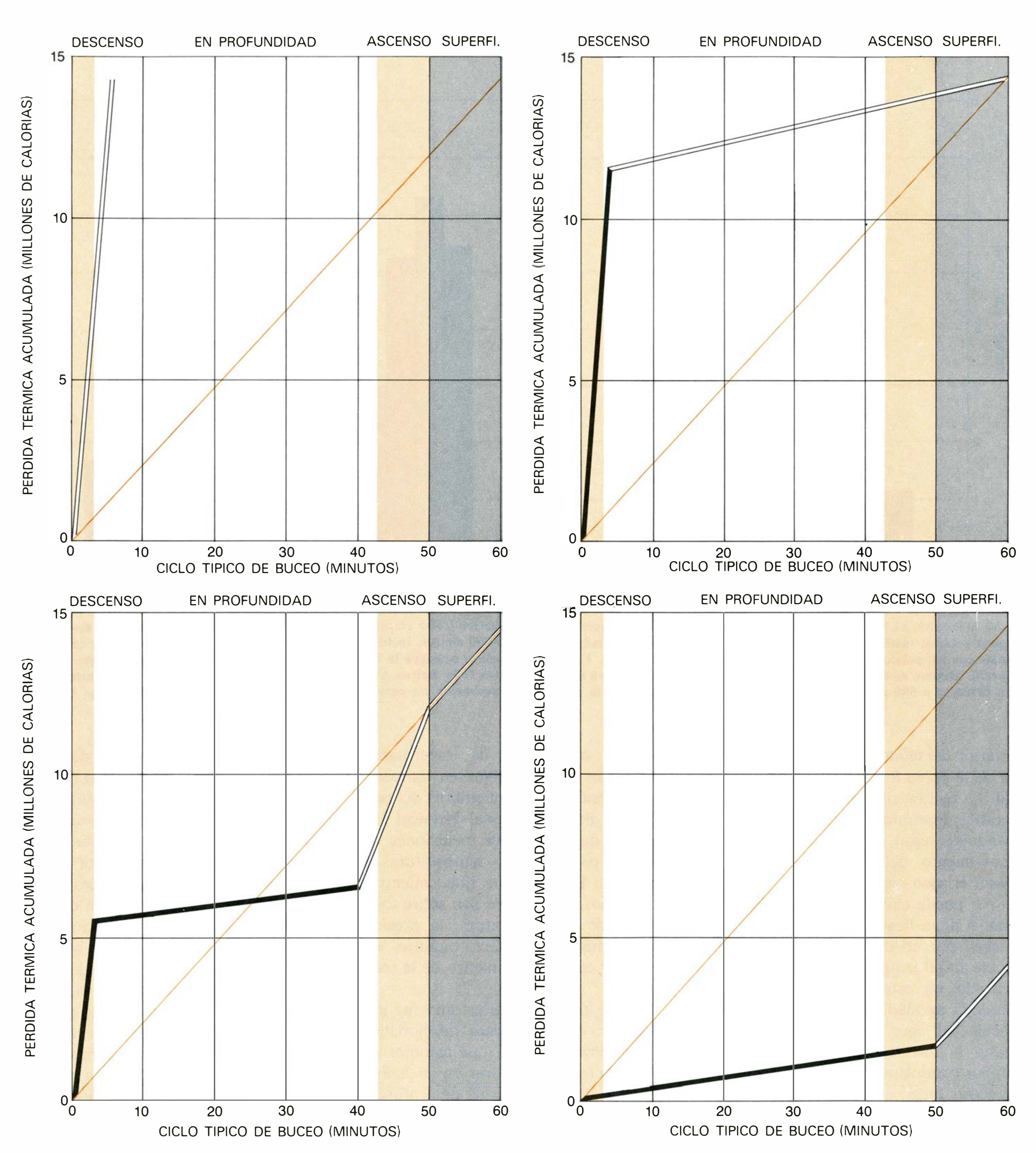
No puede excluirse la posibilidad de que el agua llegue al órgano del espermaceti por un camino distinto, el conducto nasal izquierdo. Una vez hubiera entrado en este conducto más corto desde la cavidad situada debajo del espiráculo, el agua podría ser bombeada hasta la cavidad en la que se unen ambos conductos nasales; el agua podría penetrar luego en la parte posterior del conducto nasal derecho y ser movida hacia adelante mediante la contracción de las fibras musculares del suelo del conducto. Si éste fuera el caso, cabría preguntarse por qué estas pocas fibras desempeñan un papel relativamente importante, y los grandes músculos insertos en la caja, un papel relativamente reducido. No obstante, en las dos ocasiones en que la disección reveló la presencia de agua de mar en el conducto nasal derecho de un cachalote, también encontré agua en el conducto izquierdo, de modo que no puede descartarse esta vía alternativa.

Para averiguar qué cambios en la densidad del aceite de espermaceti acompañan a los cambios en temperatura y presión, pedí a miembros del personal del British National Physical Laboratory que realizaran mediciones de densidad para diversas temperaturas y presiones. Encontraron que, mientras los cambios de densidad son sobre todo función de la temperatura, la presión tiene asimismo un efecto notorio, en particular cerca del comienzo de la congelación.

En este punto de mi estudio me enfrenté con dos problemas relacionados entre sí. Primero: ¿con qué cambios de flotabilidad se encontraría un cachalote a lo largo de toda su gama de profundidades de inmersión tanto en un ambiente polar como en uno ecuatorial, suponiendo que no tuviera ningún sistema de control de flotabilidad? Segundo: ¿basta la cantidad de aceite de espermaceti en la cabeza de un cachalote adulto para conseguir la flotabilidad neutra mediante cambios de densidad a lo largo de la misma gama de profundidades y de ambientes?

En lo que respecta al primero de estos problemas, los factores que influyen en la flotabilidad del cachalote comprenden la densidad del agua de mar, el efecto de la presión sobre los componentes líquidos y sólidos de los tejidos del cetáceo y el efecto de la presión sobre el volumen de sus pulmones. Todas las variaciones de estos factores siguen leyes físicas conocidas, por lo que puede calcularse su efecto sobre un cachalote de cualquier tamaño. Los cálculos, naturalmente, dependen de estimas bastante precisas de las proporciones de líquidos, sólidos y gases en el odontoceto en inmersión, así como de la flotabilidad del mismo cuando éste nada en la superficie. Con un animal tan grande no es fácil realizar cálculos precisos de este tipo. Los que hemos utilizado en nuestro estudio incluyen, por tanto, valores máximos, mínimos y medios; en la argumentación que sigue se utilizan los valores medios, pero las conclusiones continúan siendo válidas cuando se utilizan en su lugar los valores extremos.

Al desplazarse sobre la superficie entre latitudes ecuatoriales y polares, un cachalote experimenta un cambio en la temperatura de las aguas superficiales de hasta 26 grados C. Además, si el cetáceo se sumerge hasta una profundidad de 1000 metros en el ecuador, el cambio de temperatura puede ser de hasta 23 grados C. Estos cambios de temperatura van acompañados de variaciones en la densidad del agua de mar. La densidad depende principalmente de la tempera-



BALANCE TERMICO de un ciclo de buceo. Dicho balance exige una pérdida de cerca de 15 millones de calorías a lo largo de un período de 60 minutos, dividido en unos tres minutos para descender a una profundidad de 500 metros *(banda de color a la izquierda)*, unos 40 minutos de acecho de la presa a esta profundidad, unos siete minutos para el ascenso a la superficie *(segunda banda de color)* y 10 minutos de descanso en la superficie *(banda gris)*. Es necesaria la pérdida de calor para equilibrar la ganancia en calor corporal resultante del esfuerzo realizado durante el ciclo de buceo. Cada uno de los cuatro gráficos representa la relación entre la tasa constante de ganancia de calor *(línea diagonal de color)* y la tasa de pérdida de calor que se consigue con cada una de cuatro estrategias distintas de intercambio de calor. Los cálculos se han hecho sobre un cachalote de 30 toneladas que se sumerge hasta 500 metros de profundidad, con los pulmones llenos y en el ecuador. El trazo abierto del gráfico *A* representa el resultado de una inmersión con un intercambio continuo de calor a través de la piel del cuerpo y de la del hocico, hasta que el aceite de espermaceti es suficientemente frío para producir la flotabilidad neutra. La línea del gráfico *B* representa el resultado de una inmersión con intercambio de calor a través del conducto nasal derecho, de la piel del cuerpo y de la piel del hocico *(segmento relleno de la curva)* hasta que se consigue la flotabilidad neutra. La pérdida de calor se reduce después mediante vasoconstricción. La línea del gráfico *C* representa el resultado de una inmersión con intercambio de calor a través del conducto nasal derecho sólo durante el descenso; posteriormente, la pérdida de calor se reduce por vasoconstricción hasta que pasan unos 40 minutos *(sección abierta de la curva)*. Después el calor se transfiere desde el cuerpo al aceite de espermaceti hasta que la temperatura del aceite sube a 33 grados C. Subsecuentemente, la vasodilatación permite un intercambio térmico ulterior a través de las superficies dérmicas. Esta estrategia produciría, de forma alternada, una pérdida neta y una ganancia neta de calor hasta que el cachalote volviera a la superficie. La curva del gráfico *D* representa una inmersión en la que el intercambio de calor se minimiza mediante vasoconstricción hasta que el cachalote retorna a la superficie. La ligera pérdida de calor como consecuencia de la vasodilatación en superficie en aguas ecuatoriales *(segmento abierto de la curva)*, junto al limitado intercambio de calor durante la inmersión, produciría un balance desequilibrado; no podría volver a bucear hasta pasados 10 minutos.

tura, pero la salinidad y la presión también intervienen en los cálculos. Por ejemplo, si un cetáceo que tiene flotabilidad neutra en superficie en el ecuador fuera trasladado a una latitud antártica de 55 grados sur, el aumento de densidad del agua de superficie le proporcionaría al animal una flotabilidad incrementada, del orden de un poco más del 0,3 por cien del peso del cuerpo. El aumento de la densidad del agua que el cetáceo encontraría durante una inmersión profunda, ya fuera en el ecuador o a los 55 grados de latitud sur, aumentaría asimismo su flotabilidad. Sin embargo, debido a la mayor variación en la densidad del agua con el aumento de profundidad en el ecuador, el aumento de la flotabilidad del cetáceo resultaría aquí unas cinco veces mayor de lo que sería en el Antártico: a una profundidad de 2000 metros representaría casi el 0,5 por cien del peso corporal del animal.

¿Qué ocurre con la presión hidrostática? Su efecto sobre el contenido de agua de los tejidos del cachalote resulta ser muy leve, al igual que su efecto sobre el contenido de aceite. En lo que respecta a los componentes sólidos de los tejidos del cachalote, son casi incompresibles, de modo que su masa proporciona al cetáceo un efecto ascensional adicional al aumentar la presión. A 1000 metros de profundidad, este efecto ascensional equivaldría al 0,1 por cien del peso del cuerpo, aproximadamente. Si se suman los efectos de la presión hidrostática sobre los tres componentes tisulares, el cachalote presentaría, a 1000 metros, una fuerza ascensional extra equivalente al 0,07 por cien de su peso.

La presión hidrostática actúa también sobre el gas de los pulmones del cachalote durante la inmersión, de manera que afecta a la flotabilidad de éste, en particular a profundidades de hasta 200 metros. Cuando un cachalote recién muerto se halla flotando en la superficie en los trópicos, el volumen de su cuerpo que asoma por encima del agua iguala aproximadamente al volumen de aire en sus pulmones. Por la misma razón, cuando un cachalote espira en la superficie, está cerca de alcanzar la flotabilidad neutra. A veces los cachalotes espiran antes de bucear, y en estas circunstancias la reducción de la flotabilidad debe facilitar la inmersión. En cambio, si el cetáceo se sumerge con los pulmones llenos, para cuando llega a los 200 metros el efecto de la presión hidrostática sobre el gas de los pulmones habrá producido aproximadamente la misma reducción de la flotabilidad que habría conseguido una espiración en superficie. Por debajo de los 200 metros no hay gran diferencia entre bucear con los pulmones llenos o hacerlo con los pulmones vacíos.

Si se consideran los efectos combinados de la densidad del agua de mar y de la presión hidrostática sobre el cachalote, puede calcularse el empuje ascensional o el empuje hacia abajo que el animal experimentará a distintas profundidades y en ambientes distintos. Por ejemplo, el cetáceo presenta una clara flotabilidad positiva en superficie en el ecuador. Si el animal se sumerge con los pulmones llenos, la rápida disminución del volumen pulmonar ocasiona un empuje hacia abajo igual aproximadamente al 0,1 por cien de su peso corporal a los 100 metros. En el momento en que el cachalote alcanza los 200 metros, el cambio en la densidad del agua contrarrestará en exceso este primer efecto, y el animal presentará un empuje ascensional positivo. Este empuje ascensional irá aumentando a medida que el animal descienda, hasta que hacia los 2000 metros será aproximadamente igual al 0,2 por cien del peso del cuerpo. En las latitudes antárticas, y puesto que las frías aguas superficiales son más densas que las cálidas aguas superficiales del ecuador, el cachalote inicia su inmersión con una flotabilidad mayor, de modo que no sufre un empuje hacia abajo muy intenso en la primera parte de su descenso.

Los cálculos indican que en toda su área de distribución geográfica el cachalote debe ajustar su densidad total en alrededor del 0,2 por cien de su peso corporal para conseguir la flotabilidad neutra por debajo de los 200 metros. He sugerido que podría conseguirse un ajuste de este tipo mediante la "congelación" del aceite de espermaceti del cetáceo, y también he descrito los intercambiadores de calor de que dispone el animal para enfriar el aceite almacenado hasta convertirlo en un sólido. Pero, ¿es adecuado para esta tarea el efecto de congelación? Entre los hechos conocidos que son útiles para contestar esta pregunta, se encuentran los hallazgos del National Physical Laboratory con respecto a la densidad del aceite de espermaceti a distintas temperaturas y presiones. Se sabe también que un cachalote de 30 toneladas tiene alrededor de 2,5 toneladas de aceite de espermaceti en su cabeza, y que en latitudes ecuatoriales en superficie este aceite se halla a una temperatura de 33 grados C.

Partiendo de estos hechos, podemos proceder al cálculo de la temperatura del aceite de espermaceti necesaria para contrarrestar el efecto ascensional natural del cetáceo a profundidades crecientes. Los cálculos demuestran que, por debajo de los 200 metros, la temperatura del aceite sólo ha de reducirse en unos cuantos grados (y nunca por debajo de los 29 grados C) para conseguir la densidad requerida. Por ejemplo, si un cachalote situado en latitudes antárticas espira antes de sumergirse, no es necesario que la temperatura del aceite de espermaceti descienda por debajo de los 30 grados para contrarrestar el empuje ascensional natural del odontoceto. Cuando se dibujan sobre un gráfico temperatura y densidad juntas, las pendientes de las líneas de temperatura y las de las líneas que indican las densidades necesarias son sorprendentemente parecidas.

Con los datos anatómicos y físicos a mano, puede considerarse la manera en que la disipación y la recuperación del calor han de sincronizarse durante el ciclo de buceo de un cachalote. Para empezar con un ejemplo en el ambiente ecuatorial, se encuentra que el calor generado durante un ciclo de buceo no puede disiparse probablemente durante los 10 minutos de cada ciclo que el cachalote pasa en o cerca de la superficie. La razón es que, a cualquier profundidad inferior a los 100 metros, la diferencia entre la temperatura del agua y la temperatura subcutánea del cetáceo es inferior a 2,6 grados C. Aún en el caso de que la pérdida activa de calor en superficie por vasodilatación (es decir, la expansión de los vasos sanguíneos en la piel) se combine con la pérdida pasiva de calor en profundidad por conducción, únicamente se eliminaría una parte del calor que el cachalote genera durante los 50 minutos activos de su ciclo de buceo. No puede retenerse el calor mediante vasoconstricción durante el período de inmersión del ciclo de buceo, porque parte del tiempo que pasa el cachalote en aguas profundas y frías debe dedicarlo a perder calor y a conseguir la flotabilidad neutra.

Observándolo con más atención, resulta que la mayor parte del calor que el cetáceo genera durante la parte activa del ciclo del buceo puede perderse ya sea lentamente, mediante una vasodilatación parcial controlada que funcione al ritmo de la producción de calor, ya sea rápidamente mediante una vasodilatación más extensa durante uno o más períodos cortos en el transcurso del ciclo de buceo. Una vasodilatación breve, aunque extensa, no necesitaría un control prolongado, y al mismo tiempo permitiría que el cachalote utilizara las variaciones en la densidad del aceite de espermaceti para el control de la flotabilidad. Por ejemplo, mientras el cachalote desciende, el aceite se enfriaría mediante un intercambio de calor no sólo a través

de la piel, sino también, probablemente, a través del conducto nasal derecho. Cuando el cachalote hubiera alcanzado la flotabilidad neutra, poco depués de llegar a la profundidad deseada, cualquier enfriamiento activo ulterior de su aceite se impediría por vasoconstricción. En lo sucesivo, se podrían realizar ajustes menores de flotabilidad admitiendo pequeñas cantidades de agua de mar en el conducto nasal derecho; ello eliminaría toda dependencia de la vasodilatación dérmica para perder calor.

Durante la mayor parte de los 50 minutos del ciclo de buceo que se pasan en profundidad, el cuerpo del cachalote, y en particular su masa muscular, aumenta de temperatura. Los músculos del cetáceo no necesitan aumentar la temperatura más de dos grados C para almacenar todo el calor necesario para que el aceite de espermaceti solidificado vuelva a su temperatura fluida normal. Inmediatamente antes de que el cachalote inicie su ascenso a la superficie, aumentaría el aporte de sangre al hocico, llevando calor desde su cuerpo más cálido hasta el órgano del espermaceti. De aquí que al mismo tiempo que la temperatura del cuerpo del odontoceto desciende hasta su nivel normal, el aceite de espermaceti que se está licuando aumenta la suya rápidamente hasta los 33 grados C, su temperatura normal. Al calentarse el aceite, disminuye su densidad y aumenta su volumen, de modo que la flotabilidad del cachalote pasa de ser neutra a ser positiva; esto ayudaría a hacer ascender al animal hacia la superficie aun en el caso de que éste se hallara exhausto. Todo el calor generado muscularmente que exceda el necesario para volver a caldear el aceite, podría perderse ahora mediante vasodilatación.

Utilizando como ejemplo un cachalote de 30 toneladas que se sumerge a una profundidad de 500 metros en aguas ecuatoriales, pueden considerarse ahora cinco alternativas con respecto al



CABEZA DE CACHALOTE a la que se está extrayendo la capa de grasa en la plataforma de despiece de un barco factoría en el Antártico. Se aprecian claramente los dientes de la mandíbula inferior.

método más eficiente de pérdida de calor. Como primera alternativa, supóngase que la vasoconstricción no tiene lugar sobre toda la superficie de la piel hasta que el cetáceo ha descendido hasta la profundidad deseada y ha conseguido la flotabilidad neutra. Durante el descenso se perderá calor por toda la superficie de la piel; la pérdida total de calor se aproximará a la ganancia total de calor resultante de la actividad muscular durante el ciclo de buceo. Cuando a esta pérdida activa de calor se le añade la pérdida pasiva a través de la grasa del cachalote, la pérdida total seguramente es superior a la ganancia total. Por ello debe eliminarse esta primera alternativa.

Supóngase, como segunda alternativa, que la vasoconstricción tiene lugar por toda la superficie de la piel del cachalote tan pronto como empieza el descenso, de modo que casi todo el enfriamiento del aceite de espermaceti es un resultado del intercambio de calor únicamente en el conducto nasal derecho. En estas circunstancias, el tiempo necesario para alcanzar la flotabilidad neutra será superior al que haría falta si simultáneamente se realizara el intercambio de calor a través de la piel. No obstante, la flotabilidad neutra se alcanza en unos seis minutos (12 por ciento del tiempo de inmersión de 50 minutos). La dilatación de los vasos sanguíneos del tejido del espermaceti, antes de que el cachalote comience a ascender, haría pasar calor de los músculos corporales al hocico hasta que la temperatura del aceite volviera a sus 33 grados C normales. Cualquier exceso de calor corporal generado durante la inmersión podría perderse después mediante vasodilatación.

En la tercera alternativa la hipótesis es de una vasoconstricción zonal. Los vasos sanguíneos de la piel del cuerpo del cachalote se contraen inmediatamente después de sumergirse éste, pero los vasos sanguíneos de la piel del hocico permanecen dilatados hasta que el aceite de espermaceti se enfría hasta el punto de flotabilidad neutra en la profundidad que sea. Esta secuencia produciría la flotabilidad neutra de manera todavía más rápida que el proceso esbozado en la segunda alternativa. Sin embargo, requiere la hipótesis adicional de que el cachalote puede dilatar o contraer los vasos sanguíneos de la piel de su hocico y de la piel de su cuerpo de manera independiente.

En la cuarta alternativa, el intercambio de calor a través del conducto nasal derecho y el intercambio de calor a través de toda la piel del cachalote tienen lugar simultáneamente. Esta secuencia conduce más rápidamente a la flotabilidad neutra que en el caso de la segunda o de la tercera alternativa. Además, la cantidad de calor corporal generado durante la inmersión que tiene que ser eliminado a través del agua de mar es menor.

La quinta alternativa, la más eficaz, implica intercambio de calor únicamente a través del conducto nasal derecho y de la piel del hocico.

Todas estas alternativas, a excepción de la primera, son plausibles. La cuarta tiene una ventaja sobre la tercera y la quinta, en el sentido de que no postula la capacidad de una vasoconstricción y una vasodilatación zonales. Sin embargo, antes de eliminar por esta razón las alternativas tercera y quinta, debe señalarse que esta dilatación zonal sería similar al rubor facial humano. Parece concebible, al menos, que los cachalotes también puedan sonrojarse.

Como prueba del intercambio de calor a través del conducto nasal derecho, en dos ocasiones he encontrado agua de mar en este conducto. Por tanto, las alternativas segunda, cuarta y quinta son probables, y la quinta, que implica a la vez el conducto nasal derecho y la piel del hocico, es la más probable de todas. A este respecto, en varios cachalotes recién muertos se encontró que el tejido del espermaceti más cercano a la superficie del hocico era ligeramente más cálido que el tejido más central. Esto sugiere que en el centro del hocico había estado actuando un sistema de enfriamiento eficiente, es decir, el intercambio de calor a través del conducto nasal derecho.

Los cambios en la densidad del aceite de espermaceti deben afectar desde luego al centro de gravedad del cachalote. Al aumentar la densidad del aceite durante el descenso del cetáceo, el centro de gravedad se desplaza hacia adelante; cuando la densidad disminuye poco antes del ascenso o durante el mismo, el centro de gravedad se desplaza hacia atrás. Estos cambios serían realmente ventajosos tanto durante el descenso como durante el ascenso, pues el cachalote suele moverse hacia arriba y hacia abajo casi verticalmente. Sin embargo, el desplazamiento del centro de gravedad significa que la orientación o asiento de popa a proa del cetáceo, para utilizar una expresión náutica, no es la misma cuando el cachalote está sumergido que cuando se halla en superficie. El principal factor en la orientación del cetáceo mientras éste se halla en la superficie no es el lastre del espermaceti, sino más bien la considerable flotabilidad que le proporcionan los pulmones llenos de aire, de manera que en estas circunstancias el papel del espermaceti

es trivial. Quizá, sin embargo, el desplazamiento hacia adelante del centro de gravedad en profundidad confiere realmente al cachalote su mejor orientación durante el tiempo del ciclo de buceo en el que el animal sumergido tiene una flotabilidad neutra y permanece inmóvil o nadando.

Los cachalotes que no llegan a los 11 metros de longitud tienen menos aceite de espermaceti en relación con su tamaño que los animales mayores. Estos cetáceos más pequeños son asimismo buceadores comparativamente menos profundos, y no migran a latitudes polares tan elevadas como los cachalotes mayores. En consecuencia, encuentran menores variaciones en la densidad del agua, y aunque tienen una menor cantidad de aceite, ésta debe bastar para compensar la menor variación de la densidad del agua.

Para cualquier animal que respire aire, cuya sangre sea caliente y que efectúe inmersiones prolongadas, el control de la flotabilidad presenta numerosas ventajas. En superficie, la flotabilidad extra permite que el cetáceo descanse. En el fondo, el éxito en la caza, que depende del silencio y de la buena audición, no se pone en peligro por la necesidad de que el animal tenga que nadar para mantenerse a un nivel determinado. Además, si el cachalote necesitara estar nadando para mantenerse al mismo nivel durante los 50 minutos de su inmersión, tendría que invertir una mayor cantidad de energía en su ciclo de buceo.

No son éstas las únicas ventajas que se derivan del control de la flotabilidad. Imagínese por un momento un cachalote hipotético que flotara menos en superficie de lo que es realmente el caso. En una parte de la gama de sus profundidades de buceo, así como en determinadas localidades geográficas, este cachalote imaginario experimentaría un empuje hacia abajo durante el ciclo de buceo. Este empuje hacia abajo es, desde luego, un riesgo para cualquier animal de respiración aérea que realiza inmersiones profundas de larga duración. Al tener no sólo una flotabilidad extra a todas las profundidades, sino también control sobre esta flotabilidad, el cachalote posee un sistema de inmersión con dispositivo de seguridad incorporado. Aun en el caso de que el animal quedara extenuado durante una inmersión profunda, todavía podría salir disparado hacia la superficie en busca de aire. Por ejemplo, un cachalote que se halla extenuado, debido quizás a haber realizado cortos arranques de natación a gran velocidad en persecución de calamares, puede dirigir el calor, que es un subproducto de su actividad muscular, al órgano del espermaceti y volver a adquirir una flotabilidad positiva. Este cachalote extenuado tiene las mismas posibilidades de volver de nuevo a la superficie que un animal que no esté cansado y que pueda nadar fácilmente hacia arriba.

¿Qué hay respecto a los métodos alternativos de control de la flotabilidad? Por ejemplo, tanto colapsando los pulmones como expulsando aire durante la inmersión se afectaría la flotabilidad. Analizada detenidamente, ninguna estratagema se muestra tan eficaz como el control mediante la densidad del aceite. El grado de colapsación de los pulmones depende por entero de la presión del agua, mientras que la flotabilidad depende también de la densidad del agua, que a su vez depende de la temperatura. Al aumentar la profundidad de buceo, estos factores actúan en direcciones opuestas; en dos puntos sólo de una inmersión profunda se igualan exactamente ambos factores para producir una flotabilidad neutra.

En cuanto a la expulsión de aire, sería un método ineficaz de controlar la flotabilidad en la mayor parte del espectro de buceo del cetáceo. Por ejemplo, por debajo de los 600 metros, incluso una expulsión mínima de aire reduciría muchísimo el volumen de aire restante en los pulmones del animal para el retorno de éste a la superficie; en algunas zonas geográficas un cachalote que hubiera expulsado aire en profundidad experimentaría un considerable empuje hacia abajo cuando nadara a través de la termoclina, una transición brusca en la temperatura del agua que se encuentra entre los 200 y los 100 metros. Además, una vez expulsado el aire se pierde. Los múltiples ajustes menores de flotabilidad a distintas profundidades que son posibles mediante el control de la densidad del aceite difícilmente podrían igualarse por medio de un sistema que depende de una reserva agotable.

La pérdida de calor por parte del aceite de espermaceti como un método de control de la flotabilidad supone todavía otra clara ventaja. En los trópicos, la temperatura del agua en superficie puede ser de menos de tres grados C por debajo de la temperatura subcutánea del cetáceo. A profundidades inferiores a 100 metros, el calor corporal generado por una inmersión profunda no podría perderse durante el período usual de descanso en superficie de 10 minutos, entre dos inmersiones consecutivas. Tanto el intercambio de calor que se halla implicado en la congelación del aceite de espermaceti durante la fase de descenso del ciclo de buceo como la transferencia de calor corporal implicada en la descongelación del aceite, antes y durante la fase de ascenso, colaboran a superar esta desventaja ecuatorial.

Pocos cetáceos, que se sepa, bucean tan profundamente o durante tanto tiempo como el cachalote. Otros dos cetáceos odontocetos (el gran calderón, *Hyperoodon*, y el cachalote enano, *Kogia*) poseen órganos del espermaceti. Ambos son buceadores profundos y de larga duración. A la vista de la complicada estructura de su hocico, parece probable que también estos animales utilicen el órgano del espermaceti para el control de la flotabilidad. Sin embargo, esta conjetura no puede probarse hasta saber mucho más acerca de las dimensiones, anatomía y hábitos de buceo de estos odontocetos.

Algunos otros cetáceos, por ejemplo el rorcual franco o común, *Balaenoptera physalus*, pueden efectuar excursiones rápidas a profundidades considerables. Las inmersiones del rorcual franco, quizás a profundidades de hasta 300 metros, raramente duran más de 10 minutos; la ballena nada continuamente y sale a la superficie a cierta distancia de su punto de inmersión. Estos cetáceos misticetos nadan, por lo general, más deprisa que los tres odontocetos que bucean a profundidad. Es evidente que no se sumergen durante períodos prolongados de tiempo o para alimentarse como hacen los tres buceadores profundos: su presa la constituye el *krill*, nombre que se da a ciertos camarones pelágicos que se encuentran en grandes cantidades en la superficie o cerca de ella.

Como ya he dicho, se han emitido muchas hipótesis con respecto a posibles funciones del órgano del espermaceti distintas del control de la flotabilidad. No obstante, ninguna de ellas explica tan adecuadamente el tamaño y la estructura del órgano. Pero una estructura anatómica tan compleja podría ciertamente cumplir más de una función. Varias de las otras funciones que se han propuesto son relativamente compatibles con la hipótesis de control de la flotabilidad que acaba de explicarse. Con todo, tantas peculiaridades del hocico del cachalote y de su carga de aceite se justifican mediante la hipótesis del control de la flotabilidad, que es difícil no aceptar ésta como la principal función del órgano. Sin embargo, la comprobación final de la hipótesis ha de aguardar a la medición de la temperatura o de la densidad del aceite del órgano del espermaceti en el curso de una inmersión profunda. Se trata de una tarea difícil y costosa, pero no imposible.

Termodinámica de la evolución biológica

Los seres vivos son unos sistemas termodinámicamente abiertos que intercambian materia y energía con el mundo exterior para adquirir y mantener estructuras que, a su vez, son susceptibles de evolucionar

David Lurié y Jorge Wagensberg

El siglo XIX contempló el apogeo de una ciencia determinista que culminó con la formalización más perfecta de la mecánica newtoniana. Esta ciencia exhibe un concepto de "orden" inspirado en la observación de los fenómenos astronómicos y proporciona una imagen armónica y preestablecida de la naturaleza. El tiempo constituye un parámetro sin dirección, todas las ecuaciones son reversibles y no existe privilegio alguno entre la noción de pasado y la noción de futuro. En este ambiente nació la termodinámica con su célebre "segundo principio", que aporta la irreversibilidad del tiempo y, con ella, una dirección bien definida para las evoluciones. La nueva ciencia fue aceptada dentro de la familia de la física como una entidad extraña y molesta. Pero a partir de entonces se han venido desarrollando nuevas disciplinas protagonizadas por el nuevo concepto del tiempo. Hoy la termodinámica se nos ofrece como una ciencia clave para la comprensión y descripción del concepto general del "cambio". En biología, la termodinámica ha encontrado no sólo un dominio inagotable de aplicación sino también una fuente de inspiración que ha sido decisiva para su propio progreso durante las últimas décadas. Las razones han sido fundamentalmente dos. Por un lado, el alto grado de complejidad que los sistemas vivos muestran a todos los niveles en sus estructuras (muy alejado ya del concepto de orden astronómico); por otro, la necesidad de un tiempo irreversible que describa la evolución e interacción de dichas estructuras.

La termodinámica interesa al problema de la vida desde sus mismos cimientos, léase primer y segundo principios, que rigen el conjunto de transformaciones fisicoquímicas que tienen lugar en el seno de los sistemas observables. El primer principio afirma la conservación de la energía total de un sistema en el transcurso de dichas transformaciones. Un sencillo ejemplo: el trabajo que mueve un automóvil más las pérdidas (en forma de calor, por ejemplo) equivale a la energía química de combustión de la gasolina liberada en el interior de los cilindros del motor. El segundo principio, en su versión original, describe la evolución de un sistema aislado (que no intercambia materia ni energía con su vecindad): existe una magnitud S, llamada "entropía", que sólo puede crecer durante el desarrollo de cualquier transformación de energía, de forma que, transcurrido un tiempo suficientemente largo, alcanza un valor máximo que caracteriza el estado final llamado de equilibrio termodinámico, estado en el que ningún proceso que altere el valor de S es posible. El segundo principio aporta, con esta primera formulación, un criterio de evolución hacia el futuro y adquiere una fuerza especial con la interpretación estadística de Boltzmann: la entropía es una medida del desorden molecular. Con esto, el segundo principio se convierte en una ley de desorganización progresiva y los sistemas que la obedecen, en contraste con los sistemas mecánicos, se olvidan de sus condiciones iniciales [*véase la ilustración de la página siguiente*].

De acuerdo con la teoría clásica del equilibrio, la evolución natural es una sucesión de estados cada vez más desordenados. El segundo principio habla de sistemas aislados y de equilibrio. Está claro que para aproximarnos a los sistemas vivos es necesario relajar estas dos fuertes limitaciones. En efecto, los sistemas vivos son, ante todo, sistemas termodinámicamente abiertos, es decir, intercambian materia y energía con su ambiente. Abrir esta posibilidad supone una situación muy diferente incluso en el mundo inanimado; existe ya la oportunidad para la existencia de estados ordenados de baja entropía a una temperatura suficientemente baja. Se trata de estructuras de equilibrio como los cristales. Esta generalización permite la descripción del crecimiento de un cristal hasta una situación estable; pero aunque más cerca, todavía está lejos de explicar cómo, a temperaturas ordinarias, las moléculas pueden organizarse hasta formar complejas estructuras que a su vez realizan las coordinadas funciones de los procesos vitales. Las estructuras de equilibrio citadas se mantienen indefinidamente sin el concurso de ningún proceso de transformación energética.

Hay que hacer, además, una nueva concesión: los sistemas abiertos deben evitar las situaciones de equilibrio. Se trata de otro aspecto evidente que los organismos muestran a todos los niveles. Basta pensar en los fuertes gradientes de energía o en los intensos gradientes de potencial químico en las células. En estas condiciones, un sistema puede llegar a una situación estable de no equilibrio llamada estado estacionario. En tales casos se produce una cierta entropía positiva d_iS que el sistema, dada su condición de abierto, puede disipar totalmente al exterior d_eS ($= -d_iS$). Mantiene de esta forma una estructura constante, ya que el balance de entropía global es nulo: $dS = d_iS + d_eS = 0$.

El gas de Knudsen es un ejemplo de sistema cerrado (no intercambia materia) pero no aislado (intercambia energía). En este caso se observa una evolución del sistema hacia un estado estacionario de no equilibrio, caracterizado por una separación de fases; es decir, ilustra cómo en una situación de no equilibrio se consigue la transición de un estado

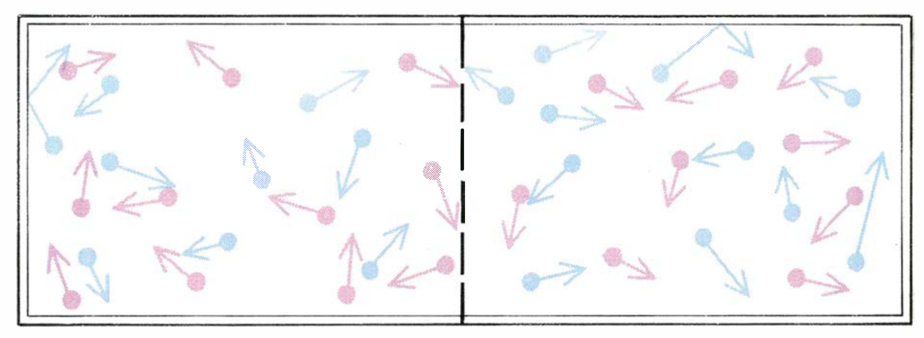
homogéneo a otro diferenciado –impuesto por las ligaduras exteriores– en el que se mantiene.

Si se aísla un sistema vivo, privándole de intercambio de materia y de energía, el segundo principio se aplica con toda su fuerza; el sistema se dirige entonces al estado de equilibrio, que es sinónimo de muerte biológica. Se necesita, pues, una termodinámica de no equilibrio para sistemas no aislados.

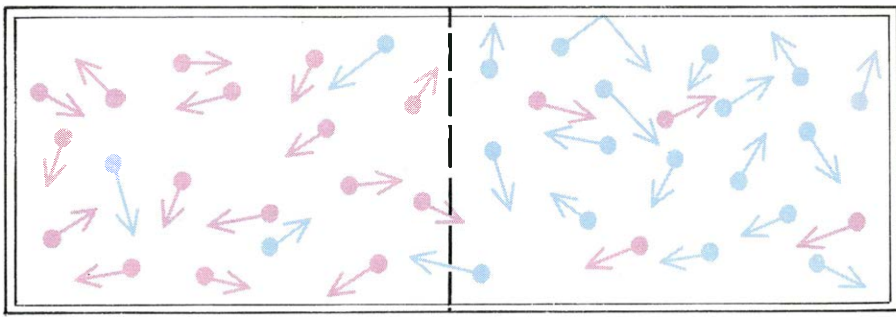
Hemos conservado hasta aquí el concepto de orden entrópico de Boltzmann y cabe preguntarse hasta qué punto puede extenderse esta idea a los sistemas vivos. [La ecuación de Boltzmann relaciona la entropía de un sistema con su grado de desorden o su probabilidad termodinámica W; es decir, $S = f(W)$.] El precedente más importante está en la obra de Erwin Schrödinger *¿Qué es la vida?* (1945). Si se admite que un aumento en el nivel de organización tiene asociado un descenso en el valor de la entropía del sistema, se obtiene un atractivo sentido biológico para el balance entrópico: la diferenciación de los organismos se consigue disipando más entropía de la que en su interior se produce. Dicho de otra manera, los sistemas vivos roban orden a su ambiente por medio de un flujo negativo de entropía, término que Schrödinger bautizara como "neguentropía". Pero, como tantas veces ocurre en un problema científico, un concepto que es satisfactorio a un determinado nivel, pierde consistencia a otros más finos. La biología es rica en un amplio espectro de niveles de observación, que van desde el molecular hasta el de poblaciones de individuos. La sospecha de que el sencillo y arquitectural orden de Boltzmann sea insuficiente para explicar el orden complejo y funcional de los procesos bioquímicos está justificada. Sin embargo, la misma termodinámica de los procesos irreversibles ofrece una nueva alternativa: el orden por fluctuaciones y el concepto de estructura disipativa. Cuando los sistemas se alejan mucho del equilibrio, la situación deja de describirse como una prolongación lógica de dicho estado. Abandonan el llamado régimen "lineal" de la termodinámica para entrar en el "no lineal". En este régimen aparecen discontinuidades e inestabilidades, el estado estacionario compatible con las condiciones que impone el ambiente ya no es único y las fluctuaciones espontáneas –antes siempre condenadas a regresar– pueden amplificarse y arrastrar los sistemas a nuevos estados estables: las estructuras disipativas. Numerosos casos de autoorganización se explican según el nuevo concepto de orden termodinámico.

La evolución natural de la termodinámica de los procesos irreversibles tiene una interesante proyección en los conceptos biológicos de crecimiento, diferenciación y evolución. Como primer paso hay que establecer el dominio de aplicabilidad de cada uno de los niveles termodinámicos a los sistemas vivos. A pesar de nuestra insistencia en el no equilibrio, importa dar cuenta de la presencia de estructuras de equilibrio en los niveles más elementales de la vida. Así, en el virus. El virus es una estructura de equilibrio cuya formación y estabilidad recuerda mucho la de ciertas estructuras cristalinas inorgánicas. Un virus aislado se mantiene como un cristal sin necesidad de ninguna clase de intercambio con el ambiente; mas para realizar una función específicamente vital, como es la reproducción, no tiene otro remedio que hospedarse en una célula, que no es sino un sistema abierto y de no equilibrio. Se trata, pues, de un caso límite. La estructura se forma fuera del equilibrio, pero es capaz de mantenerse en él.

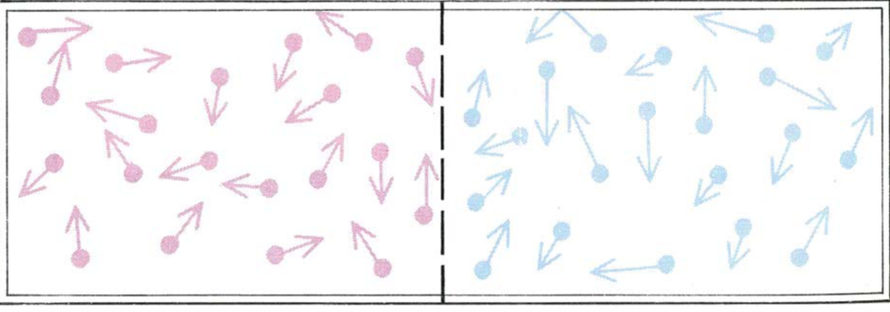
La termodinámica del no equilibrio parte del concepto de balance de entropía, introducido por Ilya Prigogine. Consideremos en primer lugar un sistema aislado. El segundo principio de la termodinámica expresa que la entropía producida en el interior del sistema es positiva durante el desarrollo de procesos irreversibles y nula en el equilibrio. ($dS = d_iS \geqslant 0$). Si el sistema no es aislado, se debe completar la expresión con un término de flujo de entropía d_eS, que representa el intercambio entrópico con el ambiente. Se establece así una ecuación de balance de entropía $dS = d_eS +$



ESTADO 1



ESTADO 2



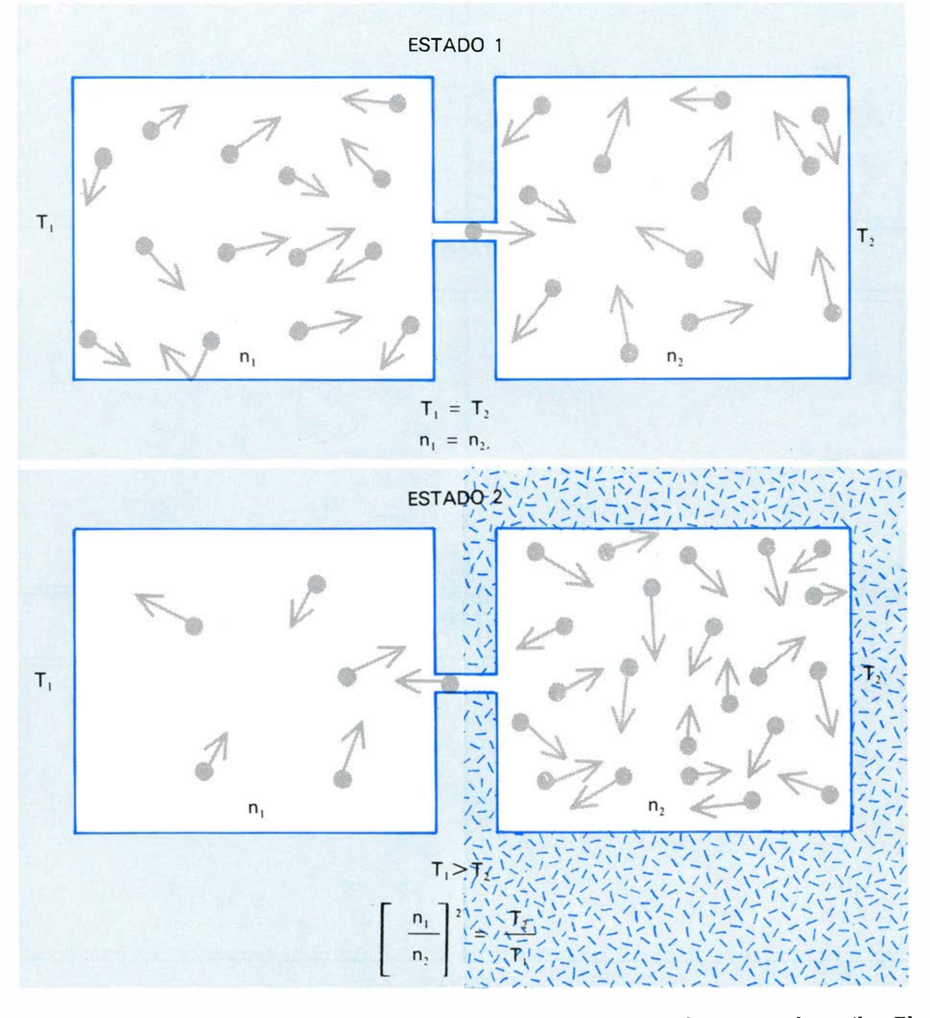
ESTADO 3

SEGUNDO PRINCIPIO DE LA TERMODINAMICA, que dice: la entropía de un sistema aislado aumenta en el transcurso de cualquier transformación físico-química. Dos gases diferentes están en un principio confinados en las dos mitades de un volumen aislado (estado 1). Si se deja evolucionar el sistema, los gases se mezclan (estado 2) y el sistema tiende a alcanzar el estado de máximo desorden (estado 3). A medida que el carácter aleatorio de la distribución de las moléculas crece, la entropía aumenta. La evolución hacia el estado 3 de equilibrio *(abajo)* es irreversible; la mezcla no se separará espontáneamente para recuperar el estado 1 inicial. El tiempo tiene aquí un sentido bien definido.

+ d_iS. Es importante que sólo el término de producción tenga un signo bien definido. Si se acepta la entropía como una medida del desorden, está claro que los sistemas aislados tienen su evolución comprometida hacia el máximo caos; pero en los sistemas abiertos existe la posibilidad de una estructuración interior si la competencia entre los términos de producción y flujo de entropía se resuelve favorablemente [*véase la ilustración inferior*].

La producción de entropía se expresa en una forma muy simple y general en términos de los llamados "fuerzas" y "flujos" termodinámicos asociados al desarrollo de un conjunto de procesos irreversibles. A cada fuerza X le corresponde un flujo conjugado J. Ejemplos de estas magnitudes correspondientes son gradiente térmico y flujo de calor, afinidad química y velocidad de las reacciones, potencial eléctrico y flujo de cargas, gradiente de densidad y flujo de masa, etcétera. La producción de entropía para un conjunto de n procesos simultáneos se escribe como la suma de los n términos formados por los productos de las fuerzas y flujos correspondientes: $d_iS/dt = \sum_{i=1}^{n} J_i X_i$. (Conviene distinguir la i subíndice de la entropía, que significa que la entropía de que hablamos es interna, de la i subíndice de fuerza, X, o flujo, J, que representa un número natural.)

En las situaciones próximas al equilibrio, flujos y fuerzas conjugados se relacionan entre sí por leyes empíricas lineales que en los ejemplos anteriores corresponden a las leyes de Fourier para la conducción de calor, de Hartley-Fick para la difusión, etc. En una superposición de procesos el flujo i-ésimo se expresa como una combinación lineal de todas las fuerzas, donde los coeficientes (llamados de Onsager) cumplen una condición de simetría: $J_i = \sum_{j=1}^{n} L_{ij}X_j$, con $L_{ij} = L_{ji}$. Esta formulación permite definir el concepto estacionario de no equilibrio y determinar los criterios de estabilidad y evolución que a él se refieren. Los estados estacionarios ocupan un lugar de interés práctico dentro de la termodinámica por su accesibilidad al tratamiento matemático y por el gran número de situaciones reales que representan. Un sistema está en un estado estacionario si sus parámetros macroscópicos tales como temperatura, presión, entropía o composición son independientes del tiempo. Esto es compatible con el desarrollo de procesos irreversibles en el interior del sistema y con que los parámetros intensivos sean función de las coordenadas especiales. La introducción de las fuerzas termodinámicas permite una clasificación muy útil de los estados estacionarios.



GAS DE KNUDSEN. Consiste en un gas encerrado en dos vasijas comunicadas por un tubo capilar. El conjunto constituye un sistema cerrado pero no aislado. Si el sistema está sumergido en un baño de temperatura uniforme, las moléculas se distribuyen de forma homogénea en el estado 1 de equilibrio. En un sistema no aislado puede establecerse un gradiente térmico fijando las temperaturas de las vasijas en $T_1 > T_2$. Se observa entonces que el sistema evoluciona hacia un estado 2 estacionario de no equilibrio que presenta una separación de fases; las moléculas que pueblan las vasijas están ahora en la relación $n_2/n_1 = \sqrt{T_1/T_2}$. Se demuestra que la entropía del estado 2 es menor que la del estado 1.

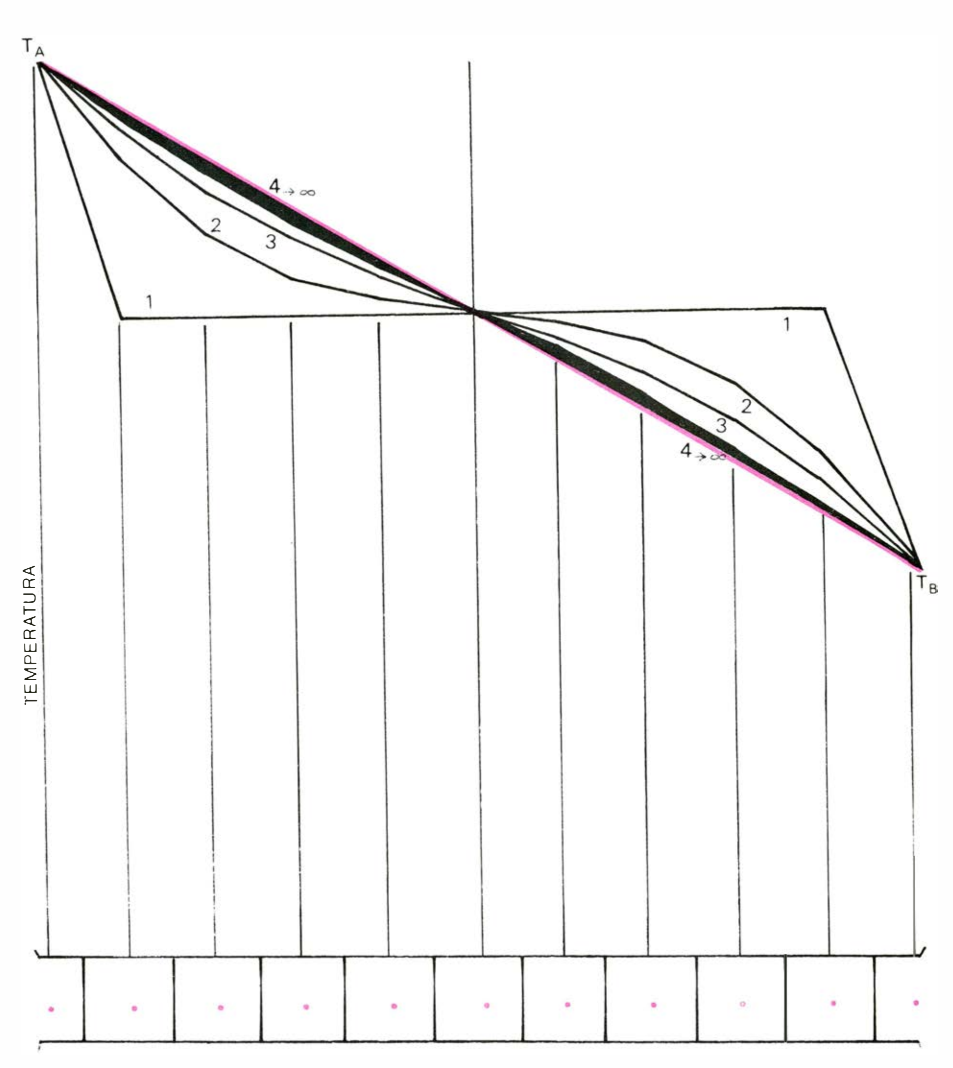
Consideremos un sistema caracterizado por n fuerzas $X_1, X_2, \ldots, X_n$, k de las cuales se mantienen fijas mediante ligaduras impuestas al sistema desde el exterior. Experimentalmente se observa que, con independencia de la configuración inicial de las n-k fuerzas libres, el sistema alcanza tarde o temprano un estado en que todas las fuerzas –y, por tanto, todos los flujos que de ellas se componen linealmente– son constantes en el tiempo. Un estado estacionario preparado de esta manera se llama estado estacionario de orden k. El equilibrio queda entonces como el estado estacionario de orden cero, o bien, como el que corresponde a no imponer ninguna ligadura termodinámica. Ocurre además que, en un estado estacionario, los flujos asociados a las fuerzas libres son no solamente constantes sino nulos, y el sistema en dicho estado queda perfectamente caracterizado por el número y la intensidad de las fuerzas fijadas. Estas fuerzas miden de alguna manera la distancia del estado al equilibrio y su debilidad garantiza la aplicabilidad de la formulación lineal.

El balance entrópico refleja en qué forma puede mantenerse un estado estacionario. El sistema envía al ambiente toda la entropía que en su interior se produce, precio ineludible de cualquier proceso irreversible según el segundo principio. Como consecuencia, el sistema mantiene su entropía constante. La persistencia de un sistema en un estado estacionario equivale entonces a la conservación de una determinada estructura o grado de organización. Por otro lado, la evolución hacia ese estado estacionario desde una configuración inicial supone la adquisición de dicha estructura final, o si se quiere, la acomodación interna del sistema a las condiciones

impuestas por el ambiente. Dentro del régimen lineal, la estabilidad del estado estacionario está asegurada; esto es, cualquier perturbación fortuita que suponga un desplazamiento del mismo es vencida por el sistema que tiende a restituir la situación estacionaria. En otras palabras, las fluctuaciones que pueden aparecer no tienen oportunidad de progresar, se amortiguan y no trascienden macroscópicamente. Se puede demostrar que esto es una consecuencia directa de la estabilidad del equilibrio. La ley de evolución hacia los estados estacionarios fue enunciada y demostrada por Prigogine; se trata del célebre principio de la mínima producción de entropía. Dicho principio establece que el estado estacionario se caracteriza por un valor mínimo de la producción de entropía compatible con las ligaduras impuestas: la producción de entropía es una magnitud no negativa que decrece durante cualquier evolución y que se hace constante y mínima una vez se ha alcanzado el estado estacionario.

Inmediatamente después del establecimiento de este principio, Prigogine y Wiame propusieron su aplicación al crecimiento y desarrollo de sistemas biológicos. Desde entonces se planteó una controversia no del todo resuelta hasta el momento. Pero veamos primero un ejemplo físico muy simple descrito por la aproximación lineal de la termodinámica: el caso de la conducción de calor en una barra metálica. Sea una barra homogénea de metal, de sección fija, longitud l y aislada en todo su entorno excepto en sus extremos *A* y *B* que están en contacto con baños a temperatura mantenida constante, T_A y T_B repectivamente. Podemos dividir longitudinalmente la barra en *n* celdas idénticas de longitud l/n cada una. La hipótesis del equilibrio local permite hacer una discretización de la distribución de temperaturas que tanto más se aproximará al caso continuo cuanto mayor sea *n*. La estructura del sistema en cada instante vendrá dada por las dos temperaturas mantenidas fijas T_A y T_B y las $n - 2$ restantes. Si $T_A = T_B$, la distribución será independiente de la posición y del tiempo, esto es, todas las celdas tendrán siempre la temperatura de los extremos. Se trata del equilibrio termodinámico como corresponde a la ausencia de ligaduras exteriores. Si T_A es mayor que T_B, el sistema evolucionará hacia un estado estacionario en que las temperaturas serán independientes del tiempo pero no de la posición.

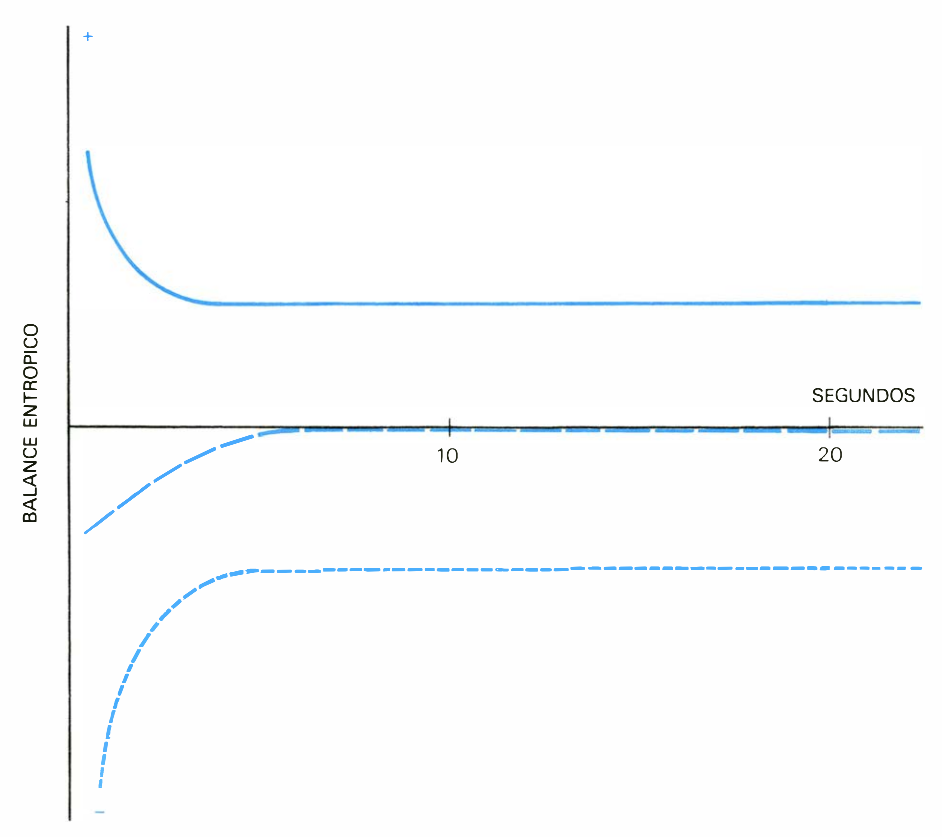
La configuración final de temperaturas y los correspondientes valores de la producción e intercambio de entropía



EVOLUCION de una configuración de temperaturas en una barra aislada en todo su contorno excepto en los extremos, donde se ha fijado la temperatura T_A y T_B. La configuración inicial 1 está representada a intervalos regulares de tiempo 2, 3, 4... ∞ (infinito), hasta alcanzar el estado estacionario (*rojo*) impuesto por las temperaturas extremas. La discretización de la barra inferior en *n* celdas iguales permite una cómoda simulación numérica de la evolución de las temperaturas y del balance entrópico.

dependen únicamente de las ligaduras termodinámicas exteriores (en este caso T_A y T_B). La evolución hacia dicho estado final de la configuración y del balance entrópico depende de la configuración inicial de temperaturas y cumple el teorema de la mínima producción de entropía [*véase ilustración superior*].

El primer problema a afrontar en el caso de los seres vivos reside en saber qué término del balance entrópico puede evaluarse experimentalmente durante el desarrollo de un organismo. Uno de los casos más discutidos es la descripción temporal del metabolismo expresado por termogramas específicos obtenidos por técnicas de calorimetría, esto es, la medida de la disipación de calor por unidad de masa y de tiempo a temperatura constante. En la ilustración de la página 107 aparece el termograma específico continuo de un sapo, según Gayda, desde el instante de la fertilización del huevo; el de una rata desde su alumbramiento, según Prat y Calvet; y la evolución del calor liberado durante el crecimiento de un cultivo de bacterias *Citrobacter* intermedium C3, calculado por unidad de célula viable, según Wagensberg y Torra. El carácter específico de estas medidas es imprescindible si se desea aproximar estos registros a los criterios termodinámicos de evolución. En la ilustración superior de la página 108 tenemos el registro térmico total de la población bacteriana en el que destacan dos fases respiratorias (aeróbica y anaeróbica) bien diferenciadas, pero que desaparecen al dividir por el número total de individuos. Tradicionalmente se ha considerado una buena aproximación el considerar dichos termogramas específicos como la producción de entropía por unidad de masa. Así lo han supuesto Prigogine, Zotin y Trintscher, entre otros. Basta una ojeada a los dos primeros casos para ver que el teorema de la mínima producción de entropía se satisface una vez se ha superado la fase de



EVOLUCION DEL BALANCE ENTROPICO correspondiente a la ilustración anterior. La curva sólida indica producción de entropía; la curva punteada, intercambio de entropía con el exterior y la curva a trazos, la variación de entropía. En el estado estacionario la entropía es constante.

embriogénesis; y que se cumple siempre en el tercer caso, donde no puede decirse que dicho período se dé en el estricto sentido de la palabra. Existen dos interpretaciones del aumento de disipación durante esta fase. Una debida a Zotin, según el cual las ligaduras no tienen por qué mantenerse constantes durante el período de embriogénesis, requisito imprescindible en la deducción del teorema de la mínima producción de entropía. Una segunda explicación más atractiva se debe a Prigogine y ha sido materializada a nivel de modelo matemático por dos investigadores de su escuela, Hiernaux y Babloyantz: durante la embriogénesis, etapa que incluye los fenómenos de diferenciación y crecimiento, los sistemas vivos escapan del equilibrio atravesando incluso la frontera del régimen lineal. Esta interpretación salva también la contradicción, ya que el principio de la mínima producción de entropía limita su validez a las situaciones próximas al equilibrio.

Según este último punto de vista los sistemas vivos utilizan procesos no lineales para aumentar su disipación al tiempo que forman estructuras disipativas espacio-temporales en los primeros pasos a través de una sucesión de inestabilidades. Y sólo una vez dicha fase se ha superado se encara una evolución hacia un estado estacionario que se caracteriza por un mínimo valor de la producción de entropía.

Aunque quizá pueda llegarse a demostrar que las dos imágenes se aproximan a la realidad, existe una tercera posibilidad igualmente sugestiva que proponemos para la interpretación termodinámica de un organismo [Lurié y Wagensberg, 1977]. Se trata de prescindir, en primer lugar, de la aproximación habitual que identifica el flujo de calor, del organismo al ambiente, con una medida de la producción de entropía y tomarlo en su acepción real de intercambio de entropía. Rechazar dicha aproximación tiene el inconveniente de que no se puede argumentar directamente sobre los termogramas utilizando el principio de evolución que afecta al término de producción y no al de intercambio. Sin embargo, permite construir una nueva imagen del desarrollo, a nivel de analogía, sin necesidad de prescindir de las hipótesis de la descripción lineal de la termodinámica. Es importante destacar que aproximar un índice experimental de intercambio a uno de producción es equivalente a anular el término de variación neta de entropía, es decir, equivale –si seguimos asignando a la entropía una cierta representación del grado de estructura– a imponer al organismo una evolución a estructura constante; y eso es tanto más cierto cuanto más cerca se está de la edad adulta, allí donde la diferenciación y el crecimiento tienden a desaparecer, y tanto menos cierto cuanto más cerca se está de la fertilización, allí donde el crecimiento y la diferenciación se imponen y donde el principio de la mínima producción de entropía parece violarse.

Recuperemos, pues, las hipótesis de las que nos serviremos para reconstruir la historia de un organismo desde la fecundación: *a)* Los datos calorimétricos suponen una información cuantitativa de la variación temporal de intercambio específico de entropía; *b)* Existen ciertas leyes fenomenológicas lineales entre fuerzas y flujos; *c)* Las ligaduras exteriores se mantienen constantes desde la fecundación.

Deben hacerse dos comentarios en relación con la tercera hipótesis. En primer lugar cabe esperar una relajación de las ligaduras exteriores superpuesta a la evolución del sistema, pero muy lenta en relación con los efectos de crecimiento y diferenciación. Bajo este punto de vista, el envejecimiento, en su sentido estricto, empieza inmediatamente y se manifiesta, una vez se ha alcanzado el estado estacionario, por el progresivo debilitamiento de las ligaduras que corresponde a la evolución gradual desde dicho estado, caracterizado por una producción de entropía positiva, hasta el de equilibrio donde aquélla se anula.

En segundo lugar, una justificación de dicha hipótesis. En la termodinámica lineal de los procesos irreversibles, un estado estacionario queda determinado –de manera única– una vez se han fijado las ligaduras exteriores. En biología, por otro lado, el estado adulto de una especie dada viene en gran parte determinado por la información contenida en el código genético. Es, por tanto, tentador suponer que, para cada organismo, las ligaduras termodinámicas exteriores quedan fijadas desde el momento de la concepción al completarse la información genética y que la degeneración del individuo se explica por una serie de errores no corregibles en el código que, como hemos dicho, suponen un olvido gradual de las ligaduras iniciales y que, en definitiva, conduce al organismo hacia el estado final de máximo desorden.

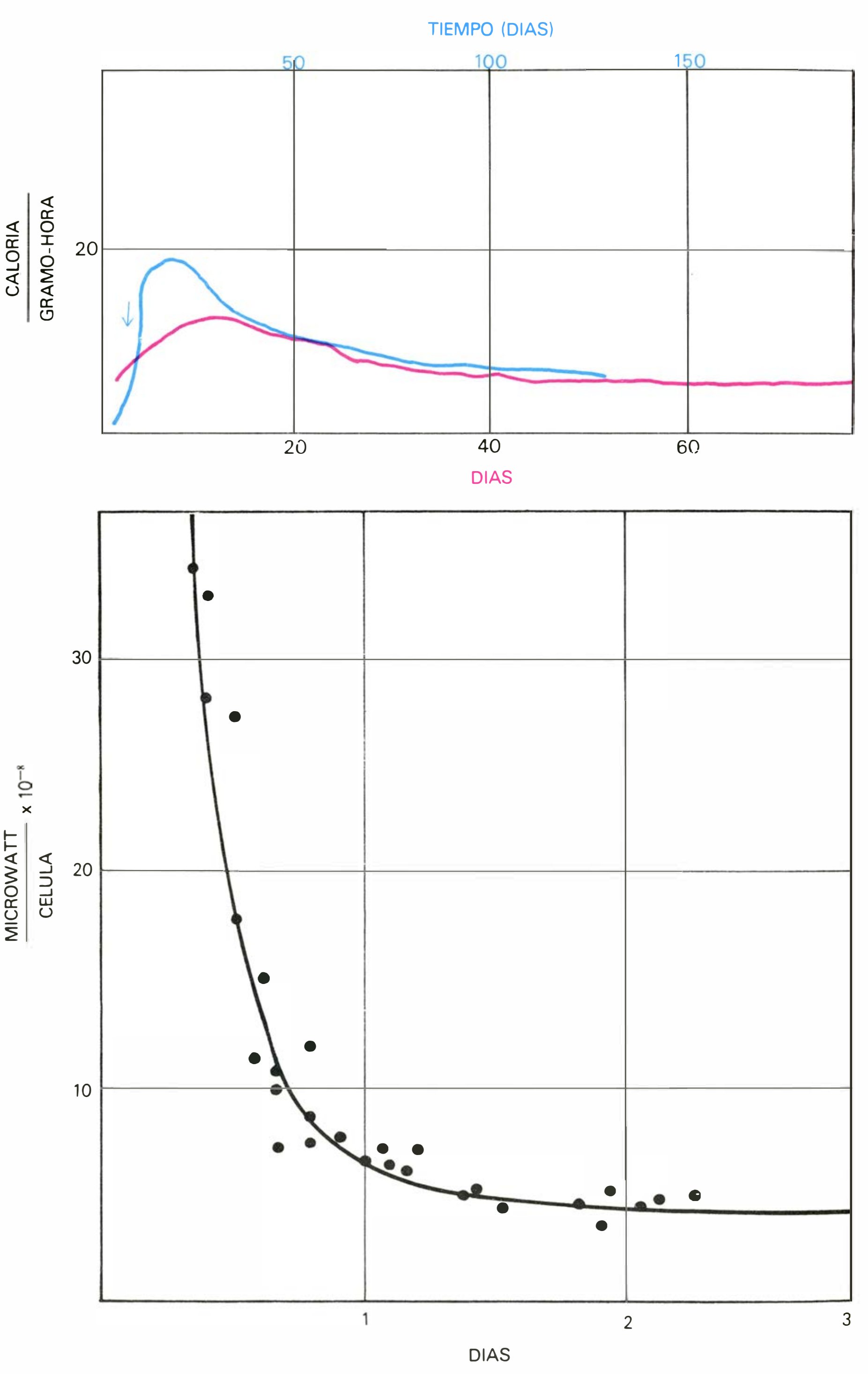
Admitiendo las tres hipótesis nos hemos armado de la siguiente información con respecto al balance de entropía que, para la evolución de un organismo, vamos a reconstruir en la ilustración su-

perior de la página 109. Por un lado, la experiencia proporciona el comportamiento cuantitativo del intercambio específico de entropía. (Se trata de un termograma generalizado del tipo de los mostrados en la ilustración de esta página; y que aparece en la ilustración de la página siguiente –curva punteada– reflejada por el eje de abscisas, para señalar que la disipación de calor corresponde a un flujo negativo de entropía para el organismo.) Por otro lado, la termodinámica propone un comportamiento cualitativo de la producción específica de entropía como una cantidad positiva y decreciente. El término resultante del balance, la variación específica de entropía, es el objetivo de nuestro análisis, ya que representa, según la sugerencia original de Schrödinger, la trayectoria estructural del organismo. Se observa que, en general, este término puede adoptar sucesivamente los signos positivo y negativo durante la evolución. ¿Cómo interpretar este hecho?

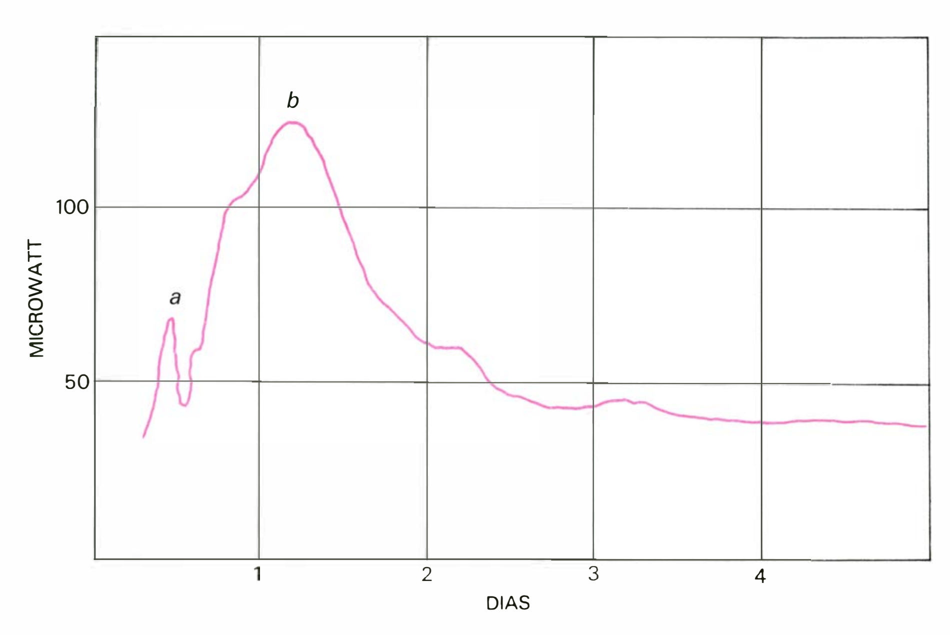
Es importante comprobar que el término de variación de entropía por unidad de masa $1/m\ dS/dt$ es el resultado de otros dos términos en competencia, con un significado biológico muy claro: crecimiento y diferenciación. El primero representa el efecto de producción de nueva biomasa y su aporte entrópico se expresa como el producto entre la tasa de crecimiento relativo $1/m\ dm/dt$ y el contenido entrópico de la unidad de masa S/m; el segundo representa la diferenciación y organización de dicha biomasa, y se expresa por la derivada temporal del contenido entrópico de masa $d(S/m)/dt$, esto es,

$$\frac{1}{m}\frac{dS}{dt} = \left(\frac{1}{m}\frac{dm}{dt}\right)\frac{S}{m} + \frac{d}{dt}\left(\frac{S}{m}\right)$$

La idea se Schrödinger se refiere, pues, al segundo término, que debe esperarse que sea negativo durante un proceso de diferenciación. El primer término es, en cambio, positivo durante un proceso de crecimiento. En definitiva, el signo del término de variación específica de entropía depende del peso relativo de los dos anteriores. A pesar de que no se dispone de un conocimiento directo de los flujos y fuerzas termodinámicos, dada la gran complejidad de los sistemas biológicos, resulta relevante el hecho de que la evolución temporal del término intercambio específico de entropía observado por vía experimental sea esencialmente el mismo para todos los organismos vivos estudiados. Esto sugiere la existencia de un denominador común relacionado con las configuraciones estructurales iniciales. En este aspecto, es difícil resistirse a la tentación de investigar la naturaleza de dichas configuraciones aunque sea recurriendo a analogías con fenómenos físicos sencillos que presenten perfiles similares en las curvas de evolución temporal de intercambio y producción de entropía. Basta atender al caso de conducción de calor en una barra metálica, comentando antes [*véase la ilustración de la página 105*], para encontrar configuraciones iniciales que proporcionen una evolución del balance entrópico en el sentido de lo que puede ocurrir en el desarrollo de un organismo vivo. Estas configuraciones iniciales resultan ser de un tipo de "estados estacio-

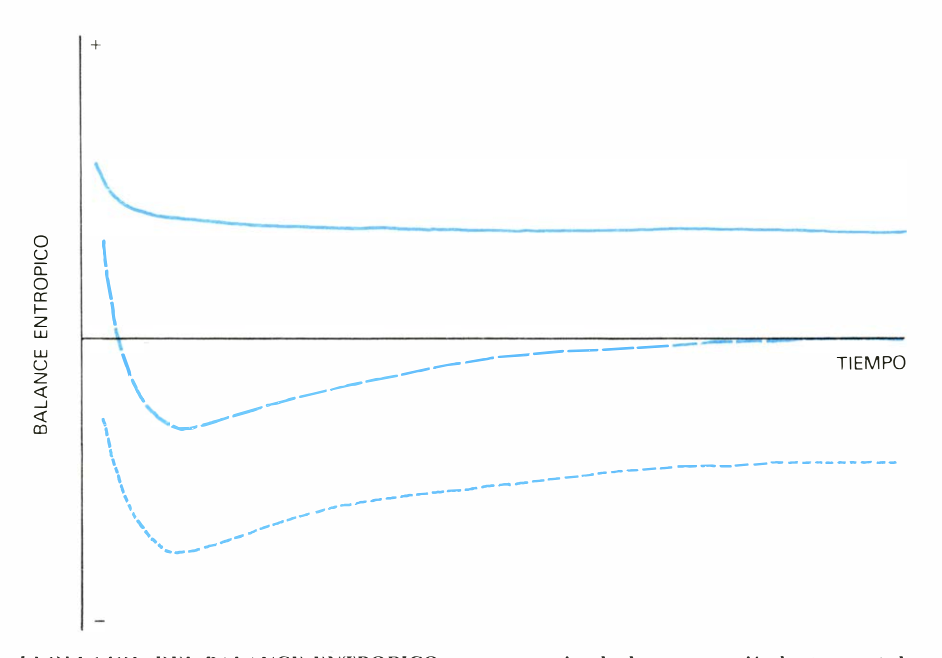


TERMOGRAMAS ESPECIFICOS de sistemas vivos. En la ilustración superior, la curva azul representa el termograma de un anfibio según Gayda desde el instante de fertilización del huevo (la flecha designa el momento de la eclosión), la curva roja representa el de una rata desde el momento de alumbramiento. En la figura inferior se registra la evolución del calor liberado durante el crecimiento de un cultivo de bacterias *Citrobacter* intermedium C 3 calculado por unidad de célula viable, según Wagensberg y Torra. Los datos se han obtenido mediante un microcalorímetro de flujo especialmente sensible.



TERMOGRAMA GLOBAL, correspondiente a toda la población del cultivo de *Citrobacter*, representada en la ilustración precedente. Se distinguen las dos fases respiratorias: aerobia (*a*) y anaerobia (*b*).

nario-perturbados" especiales que dan lugar, durante su posterior evolución, a dos efectos en competencia: uno que tiende a aumentar la entropía y otro que tiende a disminuirla. Se trata de un efecto paralelo al del conflicto crecimiento-diferenciación que sugiere la siguiente interpretación del desarrollo de los seres vivos. La fertilización transmite al sistema las ligaduras exteriores y establece su estructura inicial como una "perturbación" o "fluctuación" con respecto al estado estacionario adulto. En este aspecto, la recién creada biomasa debe contemplarse como un sistema inestable e imperfecto que requiere de la subsiguiente evolución para ir ajustándose progresivamente a las ligaduras impuestas. El estado adulto no puede mantenerse eternamente y el envejecimiento se interpreta como una aproximación gradual del estado estacionario hacia el de equilibrio por una relajación paralela de las ligaduras externas. Puede decirse que en cierto sentido el sistema se olvida de las órdenes recibidas en la fertilización a causa de una sucesión de errores moleculares que no se corrigen, es decir, de una progresiva pérdida de la información inicial.



EVOLUCION DEL BALANCE ENTROPICO para un ser vivo desde su concepción hasta su estado adulto, compatible con los datos calorimétricos y con la termodinámica de los procesos irreversibles en su aproximación lineal. La curva sólida designa la producción específica de entropía (función positiva y decreciente de acuerdo con el segundo principio y con el teorema de Prigogine); la curva punteada, un intercambio de entropía específico con el exterior (según los registros experimentales de disipación de calor); y la curva a trazos describe variación específica de entropía (función que representa el proceso de crecimiento-diferenciación). La estructura inicial determina el perfil del diagrama.

La evolución biológica adquiere, bajo esta óptica, cierta coherencia en lo que a crecimiento, diferenciación e información se refiere. La aplicación de la teoría de la información a la biología arrancó con mucho entusiasmo después de los primeros trabajos de Shannon (1948); la razón era obvia: ofrecía aquello que necesitaba la biología, a saber, una teoría cuantitativa de la organización. Así, por ejemplo, una molécula de ADN extraída de una célula de mamífero tiene una capacidad de 2×10^{10} bits (unidades de información) que puede compararse al contenido informático de cien frases de la Enciclopedia Británica. De esta forma puede evaluarse la complejidad de un organismo completo. Morowitz calculó en 1955 el valor informático de las instrucciones que serían necesarias para construír una célula de *Escherichia coli* a partir de un conjunto desordenado de moléculas. Encontró 2×10^{11} bits. Este resultado coincide bien con un cálculo que hemos realizado a partir de la entropía puesta en juego para la formación de la célula, detectada por microcalorimetría y que proporciona un valor de $3{,}5 \times 10^{11}$ bits.

Horton A. Johnson ha expresado su decepción por el hecho de que la teoría de la información y biología no acaban de converger claramente, en el sentido de que aquélla no ha permitido todavía programar o interpretar experiencias concretas en el laboratorio, que es la única medida de relevancia para cualquier teoría. Para Johnson, el organismo es una consecuencia operacional, una "deducción" de una parte de la información contenida en el huevo. Esta imagen es coherente con el supuesto de unas ligaduras termodinámicas impuestas desde el tiempo cero de la concepción a través del código genético. La información original se almacena en las células en una forma redundante, lo que supone una cierta protección, pero no se puede evitar que aquélla se pierda en el transcurso del tiempo o que se produzcan errores en el material genético. La degeneración de esta información asociada a las ligaduras heredadas no puede restituirse extrayendo información del ambiente a través de un flujo neguentrópico. El sistema es en este aspecto aislado y el caos final imparable. El individuo, incapaz de reponer parte de la información, está condenado a una pérdida continua de orden. Esto se refleja

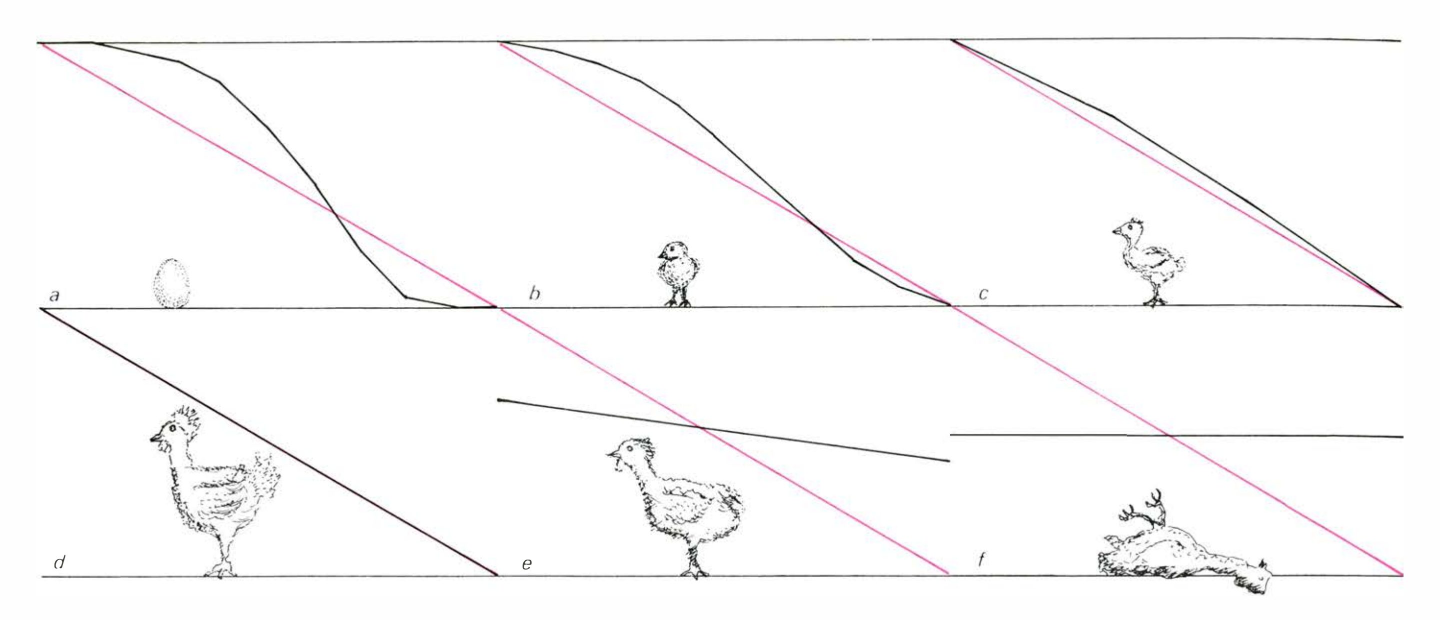


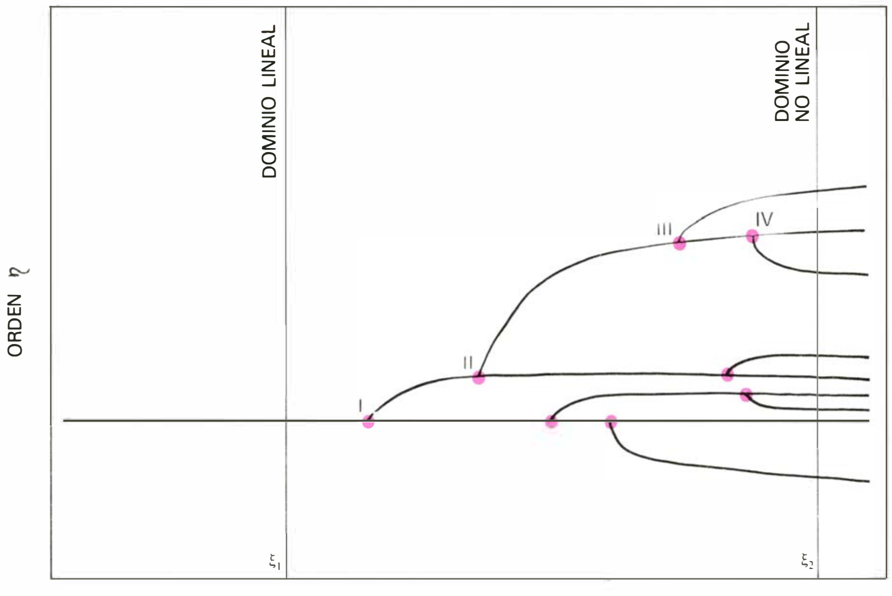
IMAGEN TERMODINAMICA de la evolución de un ser vivo (*curvas en negro*). El estado adulto (*color*) se consigue por la adaptación del sistema a las ligaduras termodinámicas transmitidas por la fecundación e impuestas por el ambiente. La degeneración de dicho estado (envejecimiento) aparece por una lenta relajación de dichas ligaduras. El estado final de muerte equivale al equilibrio termodinámico. (Ilustraciones de los propios autores.)

termodinámicamente por el hecho de que el estado adulto de un organismo no sea sino una ilusión de estado estacionario y, como hemos visto en nuestra analogía, el envejecimiento equivale a la progresiva horizontalidad de dicho estado en el diagrama hasta el fatal equilibrio final. La reproducción es la mejor solución (y la más económica desde el punto de vista entrópico) para garantizar la persistencia de la vida: hacer copias a tiempo y a bajo precio, esto es, desencadenar la formación de nuevos sistemas fuera del equilibrio por transmisión de las ligaduras pertinentes.

Hasta aquí la interpretación termodinámica –a nivel de organismo– es determinista como corresponde al dominio lineal. Quedan otros aspectos de la evolución por describir: autoorganización, mutación y selección, entre varios. Es el momento de atravesar la frontera que separa el dominio lineal del no lineal en la termodinámica de los procesos irreversibles, de dar un salto que permita descender hasta los niveles subcelular y bioquímico, que es donde se originan los acontecimientos de la evolución e interpretar fenómenos no deterministas que aparecen en niveles superiores. En esta frontera termodinámica encontramos una discontinuidad en el concepto de orden y estructura, así como en la forma en que éstos se adquieren.

Superada una distancia crítica del equilibrio, las ecuaciones cinéticas se hacen no lineales y sus soluciones pierden unicidad. Llegamos de este modo a nuevos estados de organización espacio-temporales, que Prigogine ha llamado "estructuras disipativas". Estos estados responden a un principio de orden bien diferente del principio de Boltzmann: "el principio de orden por fluctuaciones". El diagrama inferior de esta página ayuda a comprender la esencia de la nueva idea de orden sin necesidad de recurrir al complicado aparato matemático que lo describe. Sea ξ una medida de la distancia al equilibrio y η una medida de la estructura relacionada con el número de soluciones (estados estables) accesibles al sistema y compatibles con las condiciones de contorno. Para un alejamiento suficiente aparecen puntos de inestabilidad que representan una bifurcación matemática de la solución.

La multiplicidad de estados posibles se enriquece, en general, con la distancia al equilibrio. A un valor ξ_1 corresponde una solución única. Desde el punto de vista de la estabilidad del sistema, puede decirse que las fluctuaciones espontáneas, entendidas como desviaciones aleatorias de las medidas estadísticas, tienden a regresar. Su papel no llega a incidir en la evolución del sistema. Sin embargo, un valor mayor del parámetro ξ, por ejemplo $\xi_2 > \xi_1$, tiene soluciones múltiples asociadas. En este caso, las fluc-



DETERMINISMO E INDETERMINISMO según la distancia al equilibrio. En los puntos de bifurcación (*que aparecen, más gruesos, en color*), las fluctuaciones deciden el futuro del sistema.

tuaciones deciden el camino a tomar en cada bifurcación. Resulta interesante resaltar aquí la reconciliación entre azar y determinismo. La descripción de un sistema con bifurcaciones implica la coexistencia de elementos probabilísticos y deterministas: entre dos puntos de bifurcación reinan las leyes deterministas –como las ecuaciones cinéticas químicas–, pero en la vecindad de cada uno de esos puntos prevalecen las fluctuaciones. Esta colaboración entre azar y determinismo introduce un nuevo concepto de historia dentro de la física, tradicionalmente reservado a fenómenos biológicos, sociales o culturales. En efecto, llegar al estado IV del diagrama por un aumento de la disipación –y, por tanto, del parámetro ξ – conlleva el conocimiento de que las visitas a los puntos de bifurcación I, II y III se han realizado.

La formación de una estructura disipativa, auténticos nuevos estados físicos de la materia, requiere tres elementos fundamentales: *(1)* Función, esto es, la actividad fundamental descrita por ecuaciones deterministas no lineales; *(2)* estructuración espacio-temporal de dicha función, y *(3)* fluctuaciones.

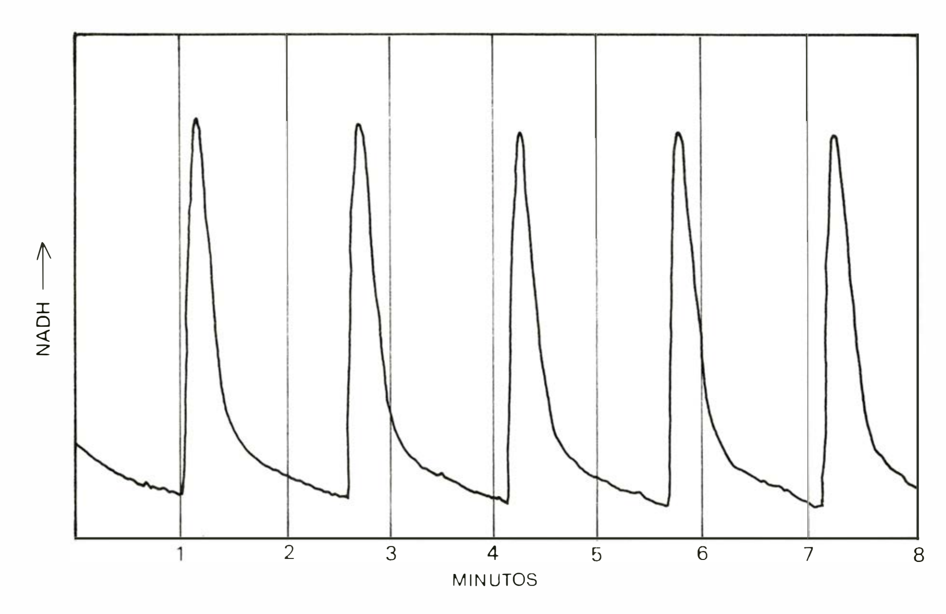
Es fácil encontrarse con situaciones experimentales y teóricas de esta clase: las reacciones autocatalíticas en química, la inestabilidad de Bénard en hidrodinámica, el láser según la interpretación de Haken en física y la teoría de Eigen para la evolución prebiótica, entre otras. La química ha sido, sin duda, un campo pionero de trabajo en este sentido, profundamente estudiado por la escuela de Bruselas. A ella pertenece el caso más espectacular de los hasta hoy descritos: la reacción de Zhabotinski-Belousov. En una solución homogénea de ciertos compuestos aparecen espontáneamente, superado un umbral crítico de disipación, unas oscilaciones alternativamente rojas y azules (excesos de iones Ce^{+3} y Ce^{+4} indicados por ferroína) que derivan gradualmente hacia una estructura en el espacio formada por estratos horizontales estables rojos y azules. Esta especie de bandera azulgrana puede mantenerse en un tubo de ensayo de 10 mililitros durante una hora, e indefinidamente en un sistema abierto en el que se establezca un flujo apropiado de reactivos [*véase la figura de la página 111*].

Recuperando en síntesis lo expuesto cabe insistir en el hecho de que dentro de la termodinámica puede concebirse la creación de estructuras en tres sentidos: de equilibrio, cerca del equilibrio y lejos de él. Pero la gran diferencia es la que separa los dos primeros del tercero. En el dominio lineal, la creación de estructuras está gobernada por el principio de orden de Boltzmann: el estado estacionario final está estrictamente determinado por el extremal (un mínimo o un máximo) de un potencial termodinámico apropiado; las fluctuaciones desempeñan un papel secundario. Lejos del equilibrio ocurre todo lo contrario, la existencia de un potencial termodinámico está comprometido y son las fluctuaciones quienes inducen nuevas estructuras amplificándose y arrastrando al sistema; rige el principio de orden por fluctuaciones. La diferencia que separa ambos principios puede resumirse cualitativamente en la propiedad siguiente. En la proximidad del equilibrio un sistema se *ajusta* a las ligaduras impuestas, con la obsesión de acomodarse en un estado lo más próximo posible al equilibrio. Lejos del equilibrio, un sistema se *reorganiza* mediante la amplificación y posterior estabilización de fluctuaciones en presencia de los intensos flujos de materia y energía que el ambiente impone. Adaptación en un caso y reestructuración en el otro.

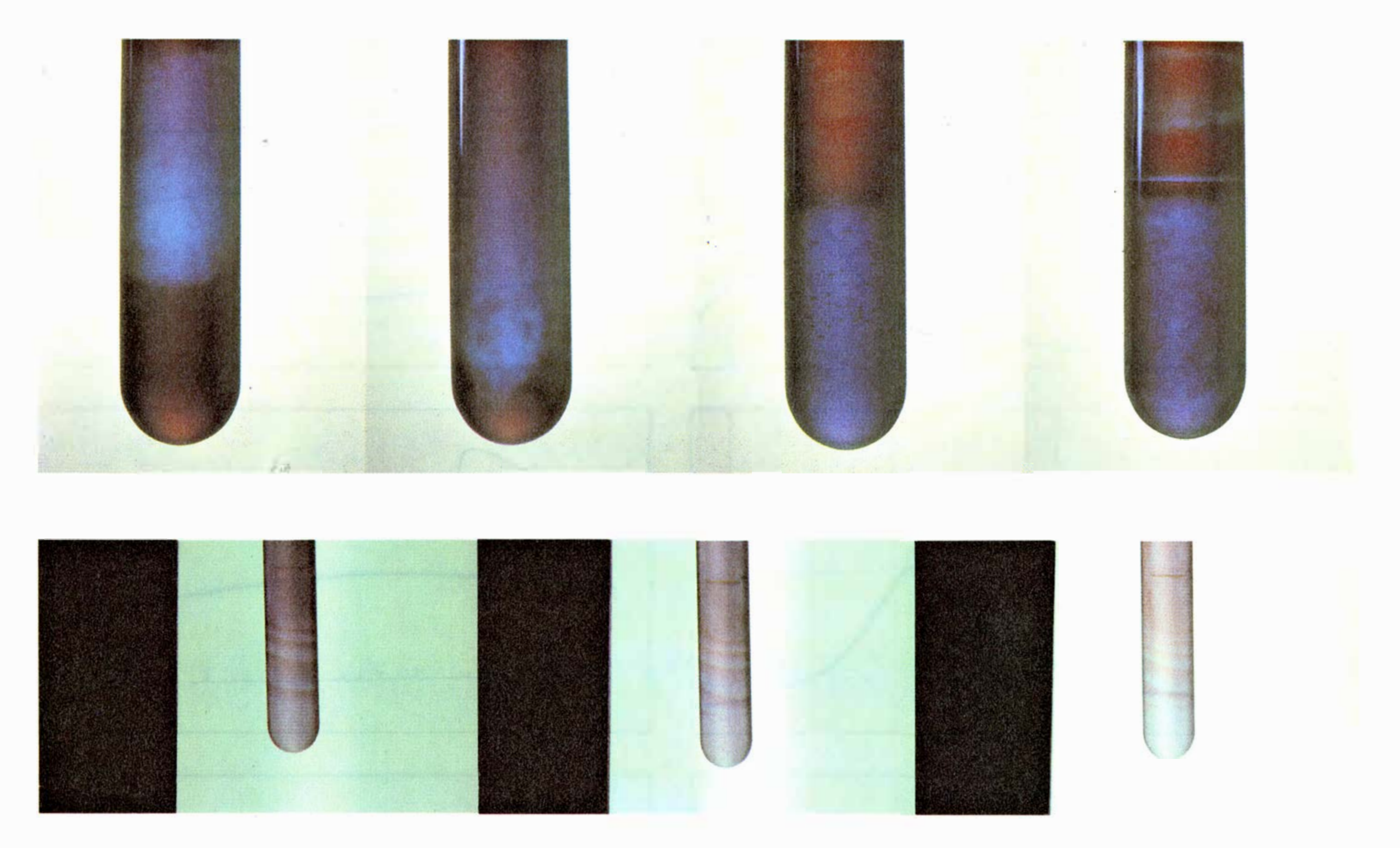
El orden por fluctuaciones plantea numerosos interrogantes en biología. ¿Cuál es su impacto en la evolución biológica y en la prebiótica? ¿Cuál es su convergencia con la teoría de Darwin? ¿Cómo se hace compatible con la imagen lineal ofrecida para el crecimiento y desarrollo a nivel de organismo? Antes de aventurar cualquier respuesta, merece hacerse una revisión rápida de las pruebas experimentales más importantes registradas y de las especulaciones propuestas a nivel de modelo.

La química del metabolismo reúne todas las características para la aparición de estructuras disipativas, léase reacciones complejas no lineales con etapas catalizadas por enzimas, fenómenos de transporte por difusión, paso de iones a través de membranas, etcétera. Una de las cadenas bioquímicas mejor estudiadas, desde este punto de vista, es la glicolisis. La experiencia muestra cómo ciertos intermediarios metabólicos presentan oscilaciones temporales no amortiguadas en su concentración, con períodos y amplitudes reproducibles. El comportamiento del NADH, por ejemplo, se visualiza por fluorimetría [*véase la ilustración de esta misma página*]. No es difícil construir modelos basados en los pasos fundamentales de dicho ciclo y en datos conocidos. Goldbeter y Lefever han demostrado con simulaciones numéricas que los resultados experimentales pueden interpretarse cuantitativamente como una estructura disipativa temporal a la que se llega tras superar ciertas inestabilidades. Cabe esperar que una gran parte de los procesos de regulación metabólica respondan a la misma interpretación termodinámica. Existen también interesantes convergencias numérico-experimentales en el control de síntesis de proteínas en las células –donde una inestabilidad separa los efectos de activación e inhibición–, en el funcionamiento del sistema nervioso –donde la diferencia de densidad de carga es la fuerza que conduce al sistema a la zona de inestabilidades–, en la agregación de seres unicelulares para formar colonias, etcétera.

La autoorganización temporal por



ORGANIZACION TEMPORAL DE NO EQUILIBRIO. Oscilaciones del intermedio metabólico de la glicolisis. NADH registradas por fluorimetría. La tasa de entrada de glucosa era de 18 mM por hora.



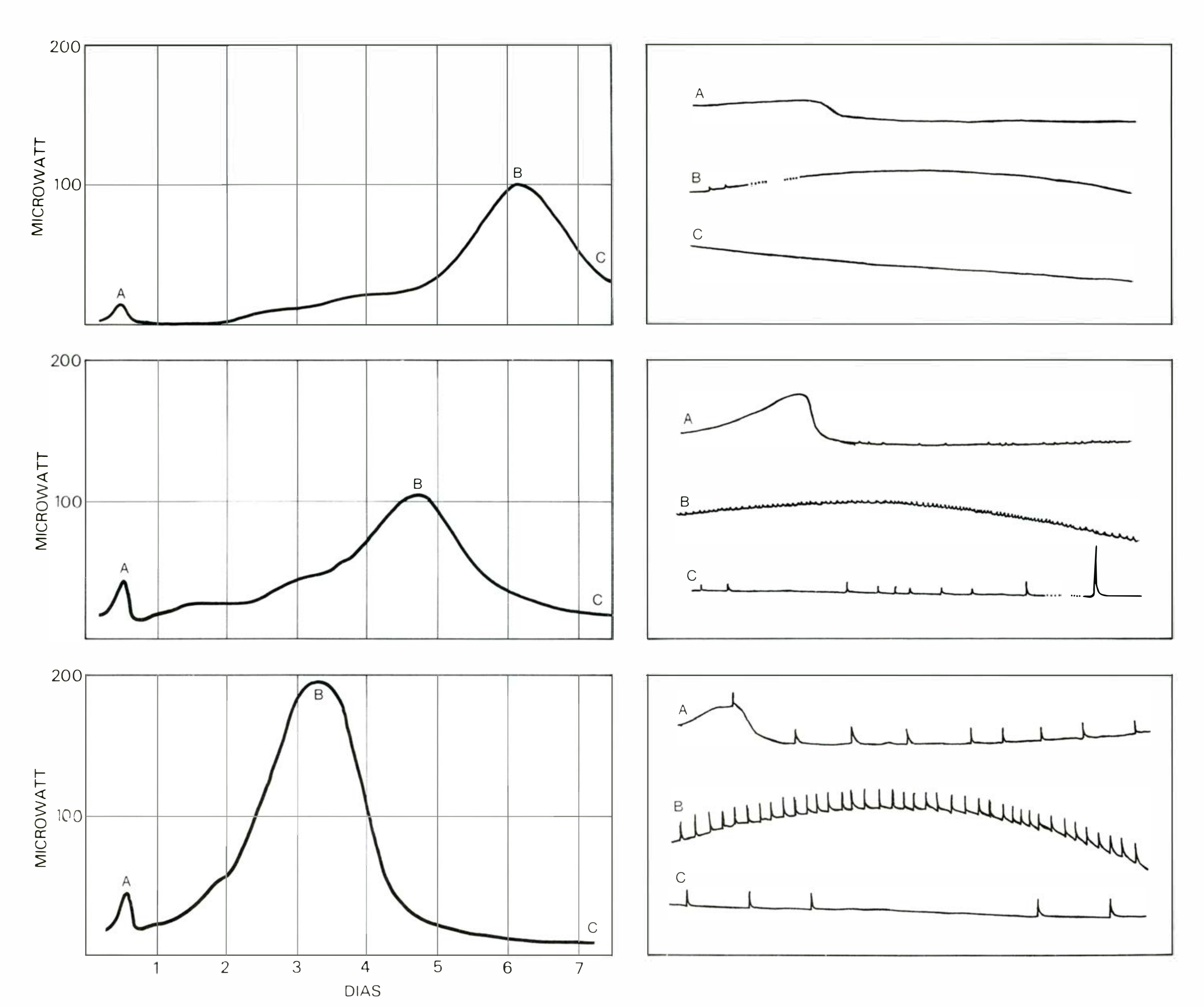
DEL CAOS AL ORDEN en la reacción de Zhabotinski-Belousov. Formación de organizaciones espaciotemporales de iones Ce^{+3} (*rojo*) y Ce^{+4} (*azul*) realizadas en el Departamento de Termología de Barcelona por los alumnos de quinto curso de Física. En la serie superior se observan oscilaciones temporales previas al asentamiento de la estructura espacial. En un sistema abierto, ésta podría mantenerse indefinidamente en el tiempo.

fluctuaciones adquiere una especial importancia en ciertos sincronismos observados a un nivel muy superior, el de poblaciones de individuos. En el sureste asiático, las luciérnagas *Luciola* se reúnen en gran número en árboles del género *Sonneratia* donde, al anochecer, emiten luz a intervalos progresivamente regulares y sincrónicos. La biofotogénesis tiene, en este caso, una motivación de convocatoria sexual y los machos consiguen con esta organización temporal enviar un mensaje más potente, que reduce el número de hembras que se despistan en el bosque.

Hemos detectado un fenómeno similar en el crecimiento de un cultivo de bacterias marinas, aisladas en el noreste de Africa por la expedición ATLOR VII del Instituto de Investigaciones Pesqueras. El estudio microcalorimétrico de dicha cepa, identificada como un *Flavobacterium*, se resume en los puntos siguientes: *(1)* La serie de termogramas obtenidos exhiben una evolución que tiende a mostrar una capacidad creciente para realizar la fase de respiración anaerobia. *(2)* Se observa un ritmo energético temporal asociado con este fenómeno. Si los termogramas se examinan localmente [*véase la ilustración de la página 112*] se puede ver cómo ciertos picos energéticos, que aparecen en la fase aerobia *A*, tienden a alcanzar una organización temporal de período constante durante la culminación de la fase anaerobia *B*. Durante la fase de muerte *C*, en la que el sistema se aproxima a un sistema termodinámicamente cerrado (dada la hermeticidad del recinto-laboratorio), la estructura temporal degenera aumentando progresivamente la distancia entre picos consecutivos. Estas oscilaciones de relajación, detectadas por calorimetría, han sido identificadas como oscilaciones glicolíticas, registrando por fluorimetría la concentración de un intermediario metabólico: el NADH. *(3)* Las poblaciones que muestran este comportamiento exhiben agregaciones de células visualmente observables, cuyo tamaño está en relación con el perfil del termograma y, por tanto, con la amplitud y período de los picos.

¿Qué sentido tiene aquí la formación de una estructura en el tiempo? ¿Tiene siquiera un sentido? En el caso de las *Luciola* la bondad de la organización resultante parece clara: mediante la sincronización aumenta la interacción entre los machos, con lo que se consigue atraer a un mayor número de hembras. Ello redunda positivamente en el futuro de la especie. La sincronización celular no está, en este momento, tan clara. Existe, sin embargo, una atractiva interpretación que servirá para describir el mecanismo por el cual el orden por fluctuaciones incide en la evolución biológica.

Un análisis de la estabilidad de las ecuaciones cinéticas no lineales del modelo alostérico de Goldbeter para el ciclo de la glicolisis, muestra que la organización de la onda química que representa al producto es especialmente sensible a la longitud del sistema. Para un valor suficientemente grande de dicha distancia (distancia supracelular) nacen ondas de concentración en el contorno del sistema que se propagan en el espacio hasta colisionar en su centro. En dicho punto, el pico resultante desaparece lentamente hasta que nuevos frentes de onda vuelven a nacer en el contorno. Este efecto resulta ser periódico en el tiempo. Puede obtenerse que en el punto central se supere, en un pequeño intervalo de tiempo, un umbral químico crítico para ciertas sustancias que son vertidas al medio extracelular en forma de pulso y cuyo calor de dilución recoge el termograma en forma de pico. Según esta imagen la unidad del nivel supracelular no conserva las características de



EL CASO DEL FLAVOBACTERIUM MARINO según Rodellar y Wagensberg. Los termogramas (*izquierda*) muestran la adaptación de esta cepa (aislada en condiciones favorables de oxígeno) para la realización de la fase anaerobia. A la derecha, tres detalles de los mismos que muestran la organización temporal asociada a la formación de agregaciones. Esta estructura del no equilibrio tiene un sentido en el campo de la evolución biológica.

una célula. La nueva unidad (una unidad de reacción-difusión "macroscópica") es un conjunto de células que oscilan en fase. Cada una de estas unidades tiene un comportamiento diferente, si bien cooperativo, con respecto al sistema global que puede considerarse como una nueva jerarquía de orden biológico: la agregación. Se han advertido casos de agregación análogos, como el del *Dyctiostelium discoideum*. El caso del *Flavobacterium* descrito contiene, sin embargo, aspectos especialmente relevantes. Por un lado, la evolución y éxito del efecto cooperativo se ha registrado en el laboratorio desde un estado "salvaje" hasta otro "educado". Por otro lado, resulta sugestivo el hecho de que estas organizaciones de no equilibrio y no lineales (solución de las leyes cinéticas) provean el mecanismo por el cual crece la interacción entre las células —inicialmente independientes— a través de la comunicación química (propagación de ondas). Esta situación podría representar perfectamente un estado de la evolución biológica intermedio entre una población de células libres (interacción débil) y una población de células con una fuerte y armoniosa interdependencia: el tejido (interacción fuerte).

En el nivel molecular y dentro de la evolución prebiótica, el concepto de orden por fluctuaciones y la teoría no lineal hallan su aplicación más importante en el modelo de Manfred Eigen. El modelo pretende describir la evolución de una población de polímeros que se supone existía ya en las condiciones de la tierra primitiva. Los polímeros tienen la propiedad de catalizar su propia replicación, pero el juego de azar puede intervenir para sustituir uno de los monómeros de la cadena por otro. Un error así se llama mutación y puede ser considerado como una fluctuación en el sentido termodinámico. Estas fluctuaciones se expresan como un término de ruido añadido a las ecuaciones cinéticas químicas que rigen las reacciones de autoreplicación. Si la mutación creada por una de esas fluctuaciones consigue catalizar su propia replicación con mayor fidelidad, tomará ventaja sobre los demás polímeros y, con el tiempo, tenderá a sustituirle por completo. Es lo que hemos llamado ampliación de una fluctuación. Las ecuaciones de Eigen aseguran, de hecho, el éxito de la población de macromoléculas que mejor se autoreplique. Se trata de un ejemplo de supervivencia darwiniana con un criterio

de selección bien definido. La evolución de la población de polímeros se visualiza como una sucesión de inestabilidades provocadas por fluctuaciones, que no son sino los errores que sobrevienen en el proceso de replicación autocatalítica. Cada estado estacionario alcanzado quedará caracterizado por la especie de polímeros dominante.

Del modelo de Eigen se desprende que un sistema formado por proteínas sin polinucleótidos nunca termina en sucesión de inestabilidades. Nunca se llega a un estado estacionario estable con respecto a los errores de replicación. Sin embargo, en un sistema mixto de proteínas y polinucleótidos, en que se pueden generar los unos con la ayuda de los otros (mecanismos de catálisis mutua), sí se alcanzan estados estacionarios de gran estabilidad con respecto a los errores. Un sistema así correspondería a un precursor del código genético.

El modelo de Eigen muestra la íntima armonía en la colaboración azar-determinismo. Las fluctuaciones (azar) tienen un papel principal: arrastran el sistema de un estado estacionario a otro. Las ecuaciones cinéticas no son menos importantes (determinismo): determinan si la fluctuación se impone o no. La evolución y la selección darwiniana aparecen así como un proceso aleatorio operando dentro de un armazón de leyes deterministas.

Esta imagen de la evolución incide en el problema del origen de la vida. Se opone al punto de vista, compartido por muchos científicos, de que la aparición de la vida se explica como una combinación de sucesos enormemente improbables, hipótesis de la que Jacques Monod es sin duda el máximo exponente. A una visión puramente indeterminista, la termodinámica del no equilibrio opone la idea de que la creación de los sistemas vivos es *inevitable* bajo las leyes de la cinética química, del transporte y de unas condiciones mínimas de alejamiento del equilibrio. Lo que no es posible es predecir en qué forma surge la vida y cuál va a ser su posterior evolución. En el caso del modelo de Eigen, por ejemplo, el estado estacionario no puede predecirse en razón del número vertiginosamente astronómico de polímeros posibles que pueden obtenerse por errores de replicación.

La noción del azar de Monod responde a una prolongación un tanto ingenua de la mentalidad de la termodinámica o mecánica estadística del equilibrio. La misma existencia de las estructuras disipativas advierte hasta qué punto es imposible extrapolar la relación orden-probabilidad desde el equilibrio hasta situaciones muy alejadas de él. La configuración de celdas hexagonales que aparece en el fenómeno de Bénard tiene, si calculamos según la mecánica estadística del equilibrio, una probabilidad infinitamente pequeña. No obstante, lo cierto es que aparece y se mantiene en condiciones mínimas de disipación. El orden por fluctuaciones y las grandes distancias del equilibrio parecen ser el soporte físico más sólido del orden biológico. ¿Dónde queda el orden de Boltzmann y las situaciones próximas al equilibrio?

El caso de la cepa marina es un buen ejemplo para poner de manifiesto la coexistencia de los dos principios, pero en niveles diferentes. Los picos comentados y el orden por fluctuaciones es un fenómeno local dentro del desarrollo de toda la población, en el sentido de que afecta a una observación de corto período. Dicho de otro modo, el mismo termograma expuesto por unidad de masa tiene un aspecto análogo al de la ilustración de la página 107, que hemos descrito dentro del régimen lineal y, por tanto, acorde con el orden de Boltzmann. Es un buen ejemplo de cómo la componente rutinaria de la evolución en un nivel está confiada a un reajuste determinista, al tiempo que el orden por fluctuaciones trabaja en una subestructura temporal. Se trata de dos comportamientos superpuestos de diferente orden de magnitud.

El orden biológico parece tener una imagen termodinámica que depende del tamaño del escenario que se observa. Un orden de Boltzmann para los sistemas grandes, donde pocas novedades pueden observarse, y un orden por fluctuaciones en los sistemas microscópicos, que continuamente aporta "ideas nuevas". La incidencia de la componente estocástica a nivel molecular garantiza a la larga lo que el ecólogo Ramón Margalef llama "lo barroco del mundo natural" cuando se refiere al hecho de que los ecosistemas contienen muchas más especies de las que harían falta si se atendiese simplemente al principio de la eficacia biológica. Huir del equilibrio predispone a la intervención del orden por fluctuaciones a modo de fábrica continua de estructuras, cuya persistencia depende de una selección posterior –en el sentido darwiniano de la palabra– según una funcionalidad medida por un criterio bien convincente: la supervivencia. En este hecho descansa la seguridad de una evolución que nunca tendrá fin.

Juegos matemáticos

Múltiples placeres matemáticos de las circunferencias mutuamente tangentes

Martin Gardner

Al estudiar las circunferencias mutuamente tangentes se suscitan centenares de hermosos problemas, muchos de los cuales se han abierto camino hasta las obras de matemática recreativa. He analizado ya en esta sección algunos ejemplos de interés histórico. Este mes estudiaré unos cuantos más. No tenemos espacio para demostraciones; los lectores que afronten el reto de encontrar demostraciones de los teoremas no dejarán de reforzar sus conocimientos y comprensión de la geometría elemental plana.

Comenzaremos por una famosa figura que, según se cree, fue estudiada por vez primera por Arquímedes, y que es conocida como "arbelos", que en griego significa "la cuchilla del zapatero", pues recuerda la forma de la hoja de una cuchilla utilizada por los remendones de la antigüedad. Se trata de la región sombreada de la ilustración inferior de esta misma página, limitada por los semicírculos de diámetro *AB*, *AC* y *BC*. *B* puede ser cualquiera de los puntos de *AC*. He aquí algunas de las más llamativas propiedades de la arbelos:

1. La longitud del arco semicircular *AC* es, como se demuestra sin dificultad, igual a la suma de los arcos *AB* y *BC*.

2. Tracemos *BD*, perpendicular a *AC*. El área de la arbelos es igual al área de un círculo de diámetro *BD*.

3. *BD* divide la arbelos en dos partes. Las circunferencias inscritas en cada una de estas partes son idénticas, y tienen cada una de ellas diámetro igual a $(AB \times BC)/AC$. La mínima circunferencia que puede circunscribir estas dos circunferencias gemelas tiene área igual a la de la arbelos.

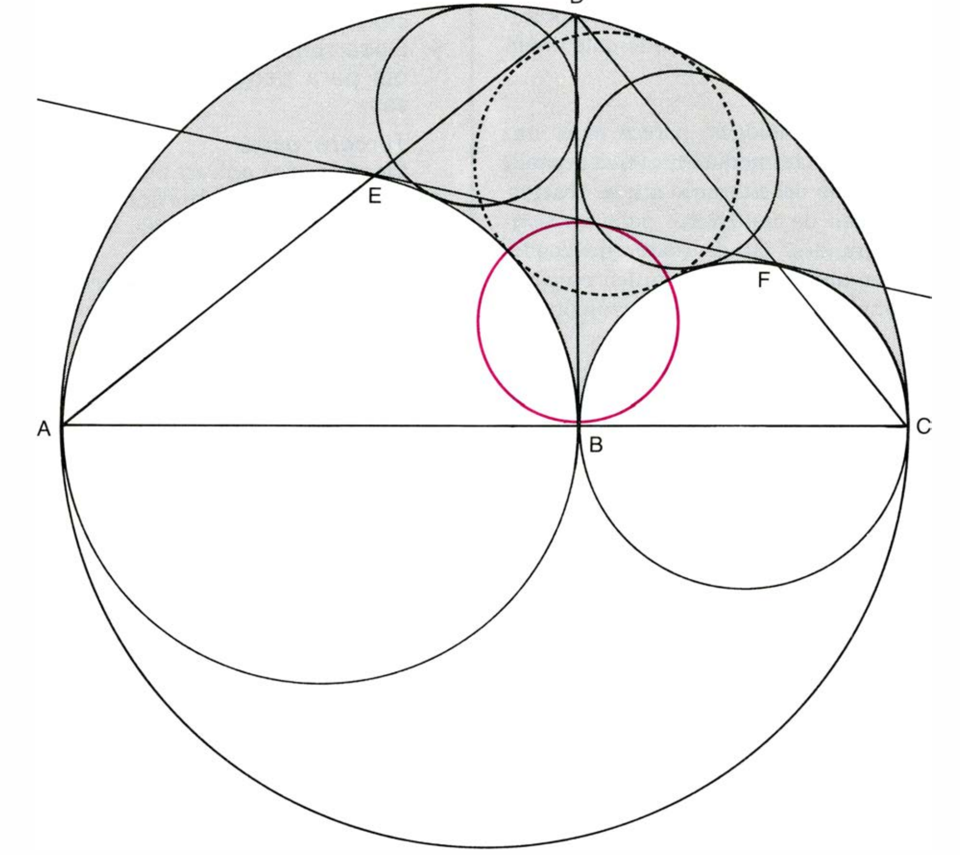
4. Trácese una recta tangente a los arcos *AB* y *BC*. Los puntos de tangencia *E* y *F* se encuentran en las rectas *AD* y *CD*.

5. Los segmentos *EF* y *BD* son iguales y se cortan en el punto medio. Esta propiedad asegura que la circunferencia de diámetro *BD* pasa por los puntos *E* y *F*.

6. A principios de la década de los 50, un matemático aficionado, Leon Bankoff, realizó un curioso descubrimiento. (Bankoff es dentista en Los Angeles, y se describe a sí mismo como "ruso por extracción". Es editor de la sección de problemas de *The Pi Mu Epsilon Journal*, y muy conocido por su habilidad para componer y resolver problemas.) Bankoff mejoró el descubrimiento, atribuido a Arquímedes, de las dos circunferencias gemelas que se ven a cada lado de *BD* en la ilustración inferior, como sigue: Se construye un círculo (cuya circunferencia se ha dibujado en línea de trazos) tangente a las tres circunferencias mayores, y seguidamente se traza una circunferencia más pequeña (en color en la ilustración) que pase por *B* y por los puntos de contacto del círculo de trazos con los arcos *AB* y *BC*. Esta circunferencia es también idéntica a las circunferencias gemelas de Arquímedes. Bankoff demuestra este resultado en su artículo "Are the Twin Circles of Archimedes Really Twins?" (*Mathematics Magazine*, vol. 47, n.º 4, septiembre, 1974, pp. 214-218). Pues no, contesta Bankoff, en realidad no son gemelas; sólo son dos circunferencias de un conjunto de trillizas.

7. Constrúyase dentro de la arbelos lo que se denomina tren de circunferencias tangentes. La circunferencia de trazos de la ilustración de la izquierda es la primera del tren, que puede prolongarse hacia la izquierda tanto como se desee, en la forma mostrada en la ilustración de la página siguiente. Marquemos C_1, C_2, C_3, y así sucesivamente, las circunferencias del tren. Los centros de todas ellas se encuentran sobre una elipse. El diámetro de cada una de las circunferencias C_n es $1/n$ -ésima parte de la distancia perpendicular desde el centro de tal circunferencia a la línea base *ABC*. Este notable resultado se encuentra en una obra del siglo IV debida a Pappus de Alejandría, quien dice, refiriéndose a él, que es un antiguo teorema.

La demostración del teorema de Pappus no es difícil recurriendo a transformaciones de inversión. Se invierte toda la figura utilizando *A* como centro de inversión; con ello, los semicírculos *AB* y *AC* se convierten en rectas paralelas, y el tren se transforma en una familia de



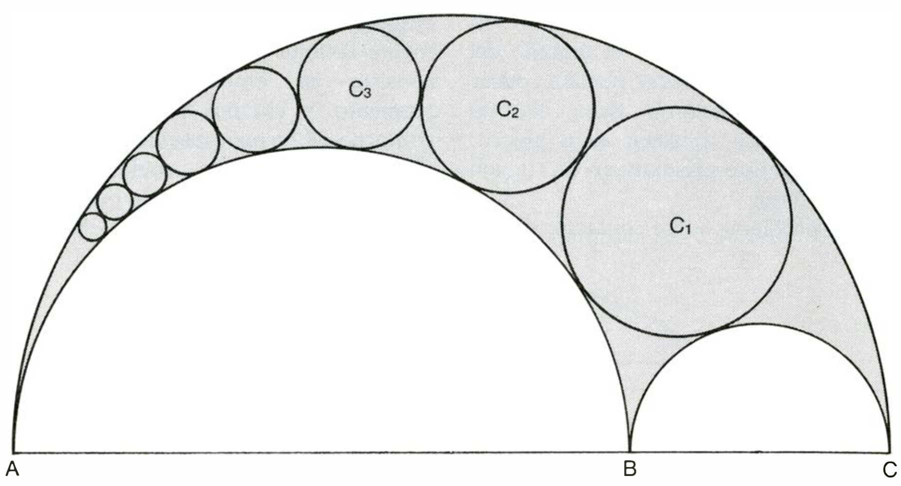
La arbelos, o cuchilla del zapatero, de Arquímedes

circunferencias iguales limitadas por estas dos rectas. Puede verse una excelente explicación de cómo se realiza todo el proceso en el artículo de Rodney T. Hood titulado "A Chain of Circles" (*The Mathematics Teacher,* vol. 54, n.º 3, marzo, 1961, pp. 134-137), y otra más breve en *Topics in Recreational Mathematics,* de J. H. Cadwell (Cambridge University Press, 1966). Pappus no conocía los métodos de inversión (que no fueron desarrollados hasta el siglo XIX), y su demostración es, por tanto, mucho más pesada.

8. Cuando AB sea 2 y AC sea 3, el tren de circunferencias ofrecerá muchas más sorpresas. Los diámetros de todas las circunferencias del tren son fracciones racionales iguales a $6/(n^2 + 6)$. Así pues, el diámetro de C_1 es de 6/7, el de C_2, 3/5, y así sucesivamente. Como me hizo observar Norman Pos recientemente, el centro de C_2 se encuentra sobre el diámetro de la circunferencia exterior trazado perpendicularmente a AC. Además, los centros de C_2 y C_3 se hallan sobre una recta paralela a AC, y lo mismo puede decirse para los centros de C_1 y C_6. Este último resultado es caso particular de un teorema general. Si AB y AC son enteros y AC es igual a $AB + 1$, los centros de cada par de circunferencias cuyos subíndices tengan producto igual a $AB \times AC$ yacen sobre una recta paralela a AC. Por tanto, si AB es igual a 3 y AC es igual a 4, los pares de circunferencias de índices 1 y 12, 2 y 6, y 3 y 4 tienen sus centros situados sobre rectas paralelas a AC. (Puede verse una demostración en M. G. Gaba: "On a Generalization of the Arbelos", en *The American Mathematical Monthly,* vol. 47, enero, 1940, pp. 19-24.)

9. Si B divide a AC en la razón áurea se verifican otras muchas propiedades sorprendentes, que Bankoff analiza en su artículo "The Golden Arbelos" (*Scripta Mathematica,* vol. 21, n.º 1, marzo, 1955, pp. 70-76). Nota para editores: Bankoff tiene inédito un original de 10 capítulos sobre la arbelos, preparado en colaboración con el matemático francés Victor Thébault.

Intimamente relacionado con la arbelos está un sorprendente teorema descubierto por Jakob Steiner, matemático suizo del siglo pasado. El teorema se describe en la ilustración superior de la página siguiente. Se traza dentro de un círculo cualquiera otro menor y se inscribe un tren de circunferencias en la región anular comprendida entre ambos. La mayoría de las veces, el tren no llegará a cerrarse exactamente –en cuyo caso formaría un anillo de circunferencias tangentes– sino que los círculos situados en los extremos se solaparán. Sin embargo, en algunos casos el tren sí formará un anillo perfecto, como el que se muestra con línea continua en la ilustración. Cuando así ocurre, se dice que el tren forma una cadena de Steiner. El descubrimiento de Steiner fue que si las dos circunferencias iniciales permiten construir una cadena de Steiner, permitirán construir infinitas más. Con más precisión: sin importar dónde se trace el primer círculo de la cadena, al añadir los demás la cadena se cierra siempre exactamente. En la ilustración se muestra también –ahora en línea de trazos– una segunda cadena arbitraria.



Tren de circunferencias en una arbelos

Como antes, la demostración más sencilla apela a métodos de la geometría de inversión. Se efectúa una inversión que transforme las dos circunferencias iniciales en circunferencias concéntricas. La cadena de Steiner se convierte entonces en una cadena de circunferencias idénticas que ocupan la corona circular definida por las concéntricas. Pueden consultarse los detalles de la demostración en la obra de Cadwell antes citada.

Solomon W. Golomb, cuyas contribuciones a la matemática recreativa aparecen frecuentemente en esta sección, tras realizar el año pasado un viaje por Europa se encontró con multitud de monedas de muy diversos tamaños. Se le ocurrió entonces la siguiente idea: Supongamos que n monedas de tamaño variable formen una cadena cerrada que rodee exactamente a una moneda central, como se muestra en la ilustración inferior de la página siguiente. Si se permuta el orden de las monedas del "corro", ¿seguirán formando las monedas un corro perfecto?

Casi todo el mundo diría que sí; incluso hay una "demostración". Desde el centro de la moneda interior se trazan rectas que pasen entre cada par de monedas adyacentes del corro, como se muestra en la ilustración. La suma de todos estos ángulos centrales tiene que ser 360 grados, hecho que parece ser independiente de la disposición de las monedas.

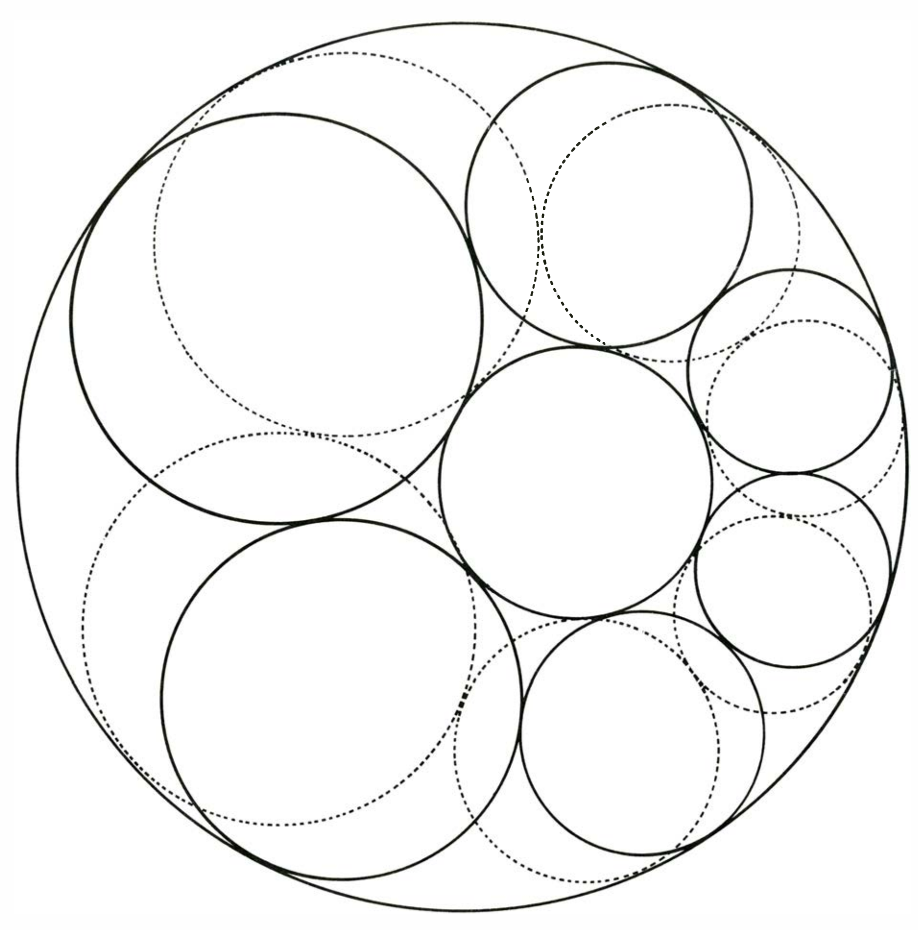
La demostración es falaz, porque, como señala Golomb, en ella se presume que las rectas que irradian desde la moneda central han de ser tangentes a cada par de monedas al pasar entre ellas. Sin embargo, no siempre ocurre así, y entonces el orden de las monedas sí tiene importancia. Evidentemente, si el corro solamente tiene tres monedas, las permutaciones carecen de importancia, pues tan sólo inducen rotaciones o simetrías de la figura inicial. Cuando haya cuatro monedas o más es fácil, según descubrió Golomb, encontrar ejemplos donde el corro se cierra con ciertas permutaciones, pero no con otras.

Las diferencias entre las distintas permutaciones son muy leves excepto cuando existan grandes discrepancias entre los tamaños de las monedas. Por consiguiente, si se rodea un círculo de tamaño adecuado con una moneda de 50 pesetas, otra de 25, otra de 5, una de 1 peseta y una de 50 céntimos, se verá que todas las permutaciones de estas cinco monedas parecen encajar exactamente. No obstante, hay diferencias. Dejo al cuidado de los lectores el trabajo de descubrir métodos de demostrar que así sucede. Quizá les guste roer otro de los descubrimientos de Golomb. Dadas n monedas, entre las que no hay dos idénticas, ¿cuál es el máximo número de círculos que las monedas pueden rodear exactamente, permutándolas a conveniencia? La solución es $(n - 1)!/2$. (El signo de exclamación representa "factorial".) Así pues, para tres monedas hay una posibilidad, para cuatro monedas hay tres permutaciones, con cinco monedas hay 12, con seis hay 60, y así sucesivamente.

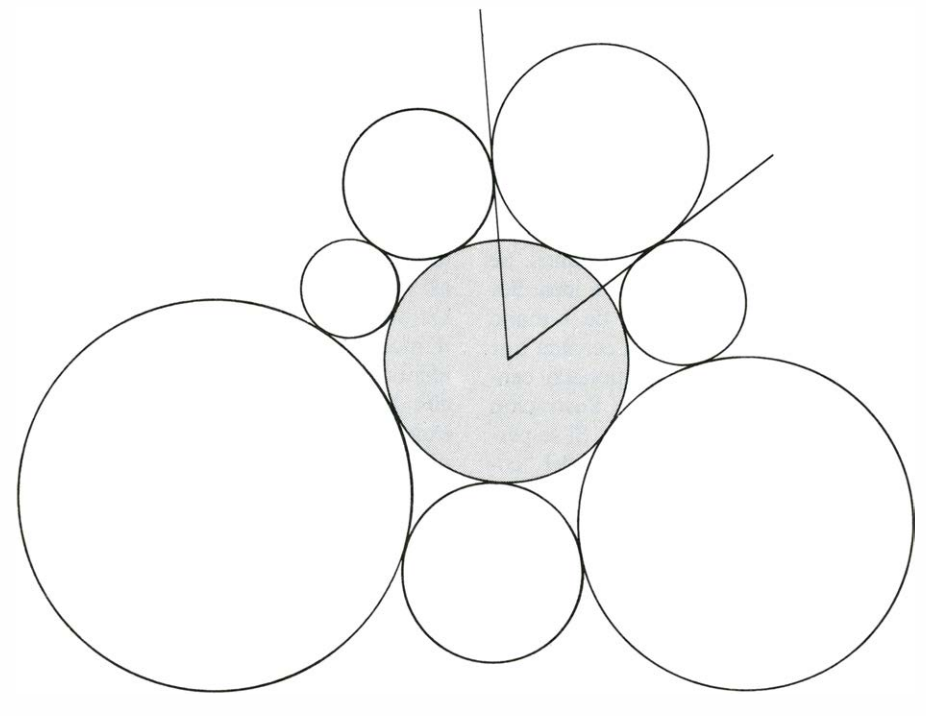
Golomb plantea también una cuestión interesante y todavía no resuelta. Dadas n monedas de distinto tamaño, ¿qué

procedimientos de colocación harán mínimo, y cuáles máximo, el tamaño del círculo que las monedas pueden rodear exactamente? Golomb tiene algunas conjeturas para construir estos algoritmos, que pronto plasmará en un trabajo sobre el tema.

Otro problema sobre contactos de circunferencias, de naturaleza completamente distinta y del que poco se sabe, consiste en empaquetar sin solapamiento n círculos idénticos en un contorno de forma especificada y de mínima superficie. Este problema no carece de importancia práctica, pues es frecuente empaquetar cilindros –latas de conserva, botellas, etcétera– en contenedores de sección transversal circular, cuadrada, o de otras formas. Dada la forma de la sección, ¿cuál habrá de ser como mínimo su superficie para que pueda alojar en ella n cilindros? Podemos enunciar el problema de forma ligeramente diferente. Dada el área de la sección, y n círculos idénticos, ¿cuál es el máximo diámetro de los círculos que consiente empaquetarlos en la región sin que se superpongan?



La "cadena" de Jakob Steiner

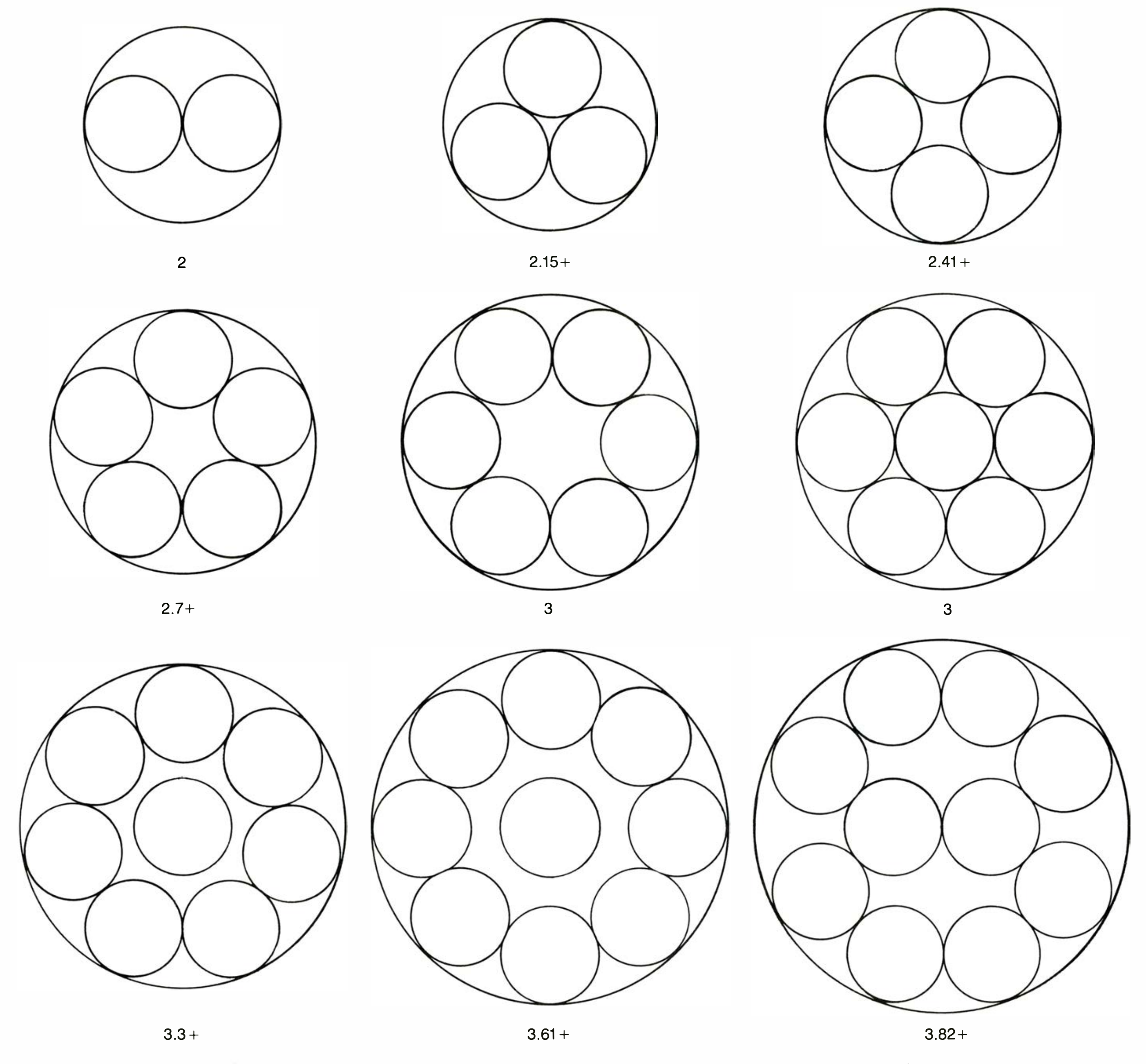


Falsa "demostración" del problema de monedas de Solomon W. Golomb

Se desconoce la solución general del problema, ni aún en casos en apariencia simples, como los de regiones de contorno circular, cuadrado, triángulo equilátero, etcétera. En cada uno de estos casos sólo han podido determinarse empaquetamientos óptimos para valores muy pequeños de n. Cuando el contorno es circular, se dispone únicamente de demostración para los casos de $n = 1$ hasta $n = 10$; fueron presentadas por vez primera por Udo Pirl, en un artículo de 1969. Los casos $n = 2$ hasta $n = 10$ (tomados de "Packing Cylinders into Cylindrical Containers", por Sidney Kravitz, *Mathematics Magazine,* vol. 40, n.º 2, marzo, 1967, pp. 65-71) se muestran en la ilustración de la página siguiente. Debajo de cada figura se da el diámetro mínimo del círculo envolvente, suponiendo siempre que los círculos pequeños sean de diámetro unidad.

Kravitz presenta también las mejores soluciones que pudo encontrar empíricamente, para $n = 11$ hasta $n = 19$. El caso $n = 12$ tiene interés especial. Podría pensarse que el denso empaquetamiento de la parte izquierda de la ilustración de la página 118 es el más denso, pero Kravitz encontró la configuración un poco mejor que se muestra a la derecha. Michael Goldberg, en "Packing of 14, 16, 17 y 20 Circles in a Circle" (*Mathematics Magazine,* vol. 44, mayo, 1971, páginas 134-139) da mejores empaquetamientos en los cuatro casos citados en el título de su artículo. A su vez, su solución para 17 circunferencias fue perfeccionada por George E. Reis (*Mathematics Magazine,* vol. 48, n.º 1, enero, 1975, pp. 33-37) quien ha conjeturado también posibles soluciones para $n = 21$ hasta $n = 25$.

Un empaquetamiento de menor mérito con $n = 12$ da, sin embargo, la solución de un popular rompecabezas llamado "El secreto del envasador", que se vendió mucho en Francia a fines del siglo pasado. El rompecabezas consistía en una caja circular que contiene 12 fichas circulares, como las damas. El problema estriba en empaquetarlas rígida y establemente en la caja, de forma que ni aún volviendo boca abajo la caja destapada caigan las fichas al suelo. En la



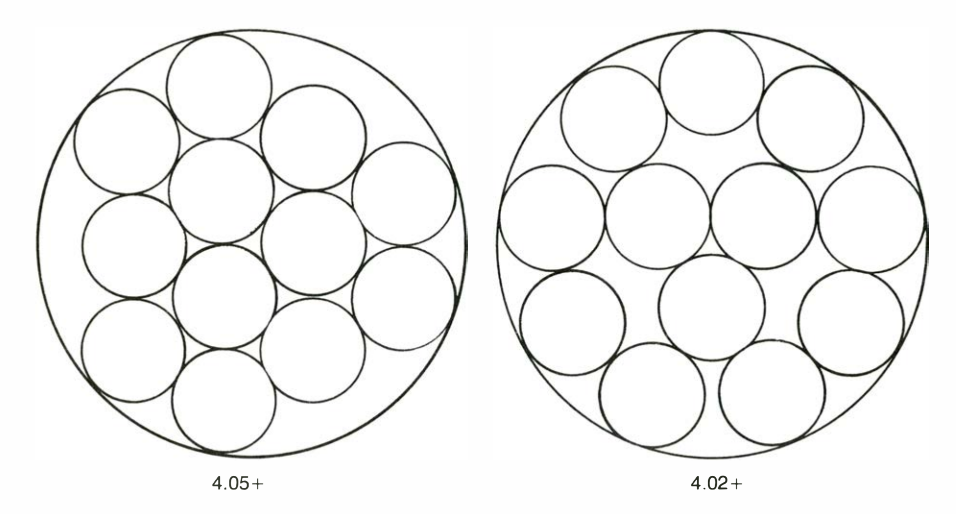
Envasamientos de máxima densidad de dos a 10 círculos unitarios en círculos

ilustración inferior de la página 118 se muestra una circunferencia que tiene el tamaño exacto para trabajar sobre el "Secreto del envasador" con 12 monedas americanas de 1 centavo. ¿Sabría el lector encajar los doce centavos dentro de este círculo, formando una configuración rígida en la que no pueda moverse ninguna moneda? Daré la solución correspondiente en el próximo número de *Investigación y Ciencia*.

Se han buscado también métodos para empaquetar n círculos iguales en un cuadrado, con densidad máxima. Se ha demostrado que al crecer n, la densidad de empaquetamiento tiende hacia 0,9069 + . Este valor es el límite correspondiente al conocido empaquetamiento denso de circunferencias obtenido situando sus centros en los vértices de una red de triángulos equiláteros. Sin embargo, sólo se conocen demostraciones que establezcan que el empaquetamiento es óptimo en los casos $n = 1$ hasta $n = 9$. Como ocurría antes, no se conoce fórmula ni algoritmo que proporcione empaquetamientos óptimos. La ilustración superior de la página 119, tomada del trabajo de Goldberg "The Packing of Equal Circles in a Square" (*Mathematics Magazine*, vol. 43, n.º 1, enero, 1970, pp. 24-30), muestra las soluciones para $n = 1$ hasta $n = 9$. En la ilustración, Goldberg muestra los máximos círculos que podrán alojarse en un cuadrado de lado unidad, en lugar de exhibir el cuadrado mínimo capaz de albergar n círculos unitarios. En su artículo, Goldberg da también las conjeturas que estima óptimas para valores de $n = 10$ hasta $n = 27$, y para algunos valores excepcionales más altos.

Goldberg demuestra que cada solución óptima ha de contener una estructura de círculos tangentes que conecte los cuatro lados del cuadrado, y que dentro de tal estructura cada circunferencia ha de establecer contacto con al menos otras tres o con uno de los lados del cuadrado. Sin embargo, pueden existir circunferencias que no formen parte de esta estructura (en el caso $n = 7$).

Bajo cada uno de los cuadrados de la ilustración puede verse el diámetro del máximo círculo posible, supuesto igual a uno el lado del cuadrado. Hay dos casos, los de $n = 6$ y $n = 7$ cuyas demostraciones no son nada sencillas, y no han sido publicadas. El primero en resolver el caso $n = 6$ fue Ronald. L. Gra-



Empaquetamiento de 12 círculos **(izquierda)**
y empaquetamiento óptimo conjeturado **(derecha)**

ham, de los Bell Laboratories. El caso $n = 7$ fue anunciado por J. Schaer en 1965. En una nota publicada en *Mathematics Magazine* (vol. 44, n.º 3, mayo, 1971, pp. 139-140), Schaer mejoró la conjetura de Goldberg para $n = 10$. Al generalizar el problema de círculos y cuadrados a esferas y cubos, su análisis se complica de un modo extraordinario, desconociéndose prácticamente todo acerca de esta versión tridimensional del problema. (Puede verse un sumario de los resultados conocidos sobre el empaquetamiento de esferas iguales en cubos en un artículo de Goldberg, en *Mathematics Magazine*, vol. 44, n.º 4, septiembre, 1971, pp. 199-208).

Los empaquetamientos de n circunferencias iguales en triángulos equiláteros presentan también difíciles problemas. Poco se ha conseguido, excepto que cuando n es número triangular (perteneciente a la sucesión 1, 3, 6, 10, 15...) se obtienen empaquetamientos de máxima densidad en disposiciones por filas de 1, 2, 3, 4, 5... círculos. Cuando el número de círculos sea igual a un número triangular disminuido en dos o más unidades, es siempre posible desplazar los círculos restantes de forma que reconstruyan un nuevo triángulo equilátero de menor tamaño. Así que, quitando dos bolas del triángulo de madera que se utiliza en el billar americano (billar de troneras) para conseguir un empaquetamiento denso de las 15 bolas al comenzar el juego, las 13 bolas restantes pueden reagruparse de forma que ninguna de ellas toque el marco.

Supongamos que solamente se retira una bola. ¿Pueden reagruparse las 14 bolas restantes de forma que ninguna de ellas toque el marco? No parece probable, pero se ignora la respuesta a ciencia cierta. Donald J. Newman ha conjeturado que en todos los casos en que el número de bolas (o círculos) sea un triangular menos uno, no hay forma posible de reagrupar las bolas y encajarlas en un marco triangular más pequeño. La conjetura no es aplicable al número triángular 1, y es evidentemente cierta para $n = 3$. Parece ser verdadera también para el siguiente número triangular, 6, pero no conozco ninguna demostración formal de imposibilidad de "estrujar" cinco bolas de forma que ocupen un triángulo menor que el necesario para alojar las seis.

Se pueden plantear cuestiones semejantes sobre empaquetamientos de máxima densidad de círculos iguales en una región de forma dada cualquiera, incluso con agujeros. En el caso de que la región carezca de agujeros y esté limitada por una curva cerrada y convexa, el mejor de los resultados obtenidos hasta ahora se debe a J. H. Folkman y a Graham, en su artículo titulado "A Packing Inequality for Compact Convex Subsets of the Plane" (*Canadian Mathematical Bulletin*, vol. 12, n.º 6, 1969, páginas 745-752). Dadas el área y la longitud del contorno de una región, los autores determinan una cota superior para el número máximo de circunferencias de diámetro unidad que pueden alojarse en ella.

Solamente tenemos espacio para mencionar uno más de los centenares de teoremas relativos a circunferencias tangentes. Se trata de un elegante resultado publicado en 1968 por el geómetra canadiense H. S. M. Coxeter. Podemos verlo en la ilustración inferior de la página siguiente. Se construye una sucesión infinita de circunferencias, de manera que cada grupo de cuatro consecutivas sean mutuamente tangentes. Resulta que la sucesión así construida es única. El radio de cada circunferencia se obtiene multiplicando el de la mínima circunferencia inmediata por la suma de la razón áurea con la raíz cuadrada de la misma, número ligeramente mayor que 2,89. Los puntos de contacto de las circunferencias se encuentran sobre una espiral equiángular, que se muestra en línea de trazos.

Círculo para trabajar sobre "El secreto del envasador"
con 12 monedas de 1 centavo de dólar U.S.A.

He aquí las soluciones de las cuestiones del mes pasado:

1. Inkstand.
2. Crankshaft.
3. Clasificando los enteros positivos en pares e impares, se tendrá que todos los pares de números de una misma clase darán suma par y mayor que 2, que por consiguiente no puede ser nú-

man, quien me comunicó este problema.)

4. El número de seguridad social pedido es 381-65-4729. Añadiendo un 0 al final se obtiene la única solución del mismo problema que utiliza los diez dígitos, del 0 al 9. (Carlos Morales Díaz, de Madrid, resolvió este problema.)

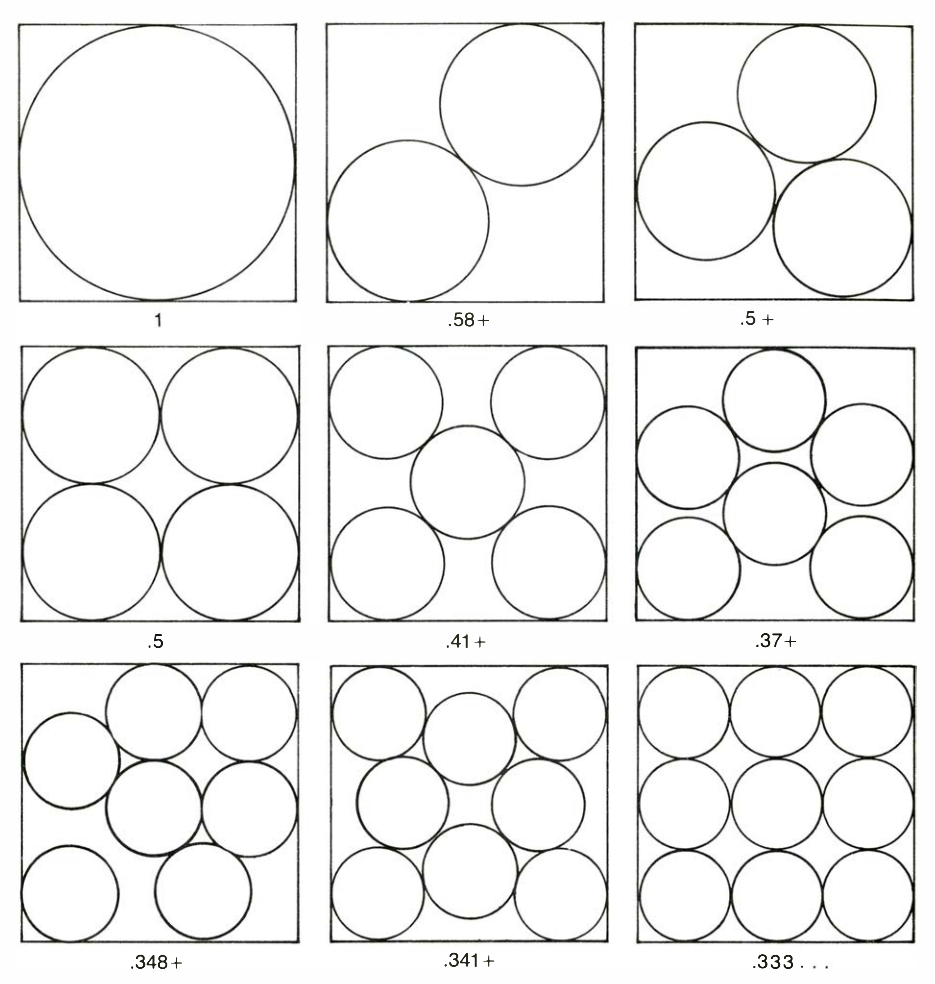
5. Se da seguidamente una solución

5. Se da seguidamente una solución en 16 movimientos del problema de intercambiar las posiciones de los caballos. Lo mismo que en la ilustración del mes pasado, se han numerado las filas de 1 hasta 4, contando de abajo arriba, y las columnas se han marcado *A*, *B* y *C*, yendo de izquierda a derecha.

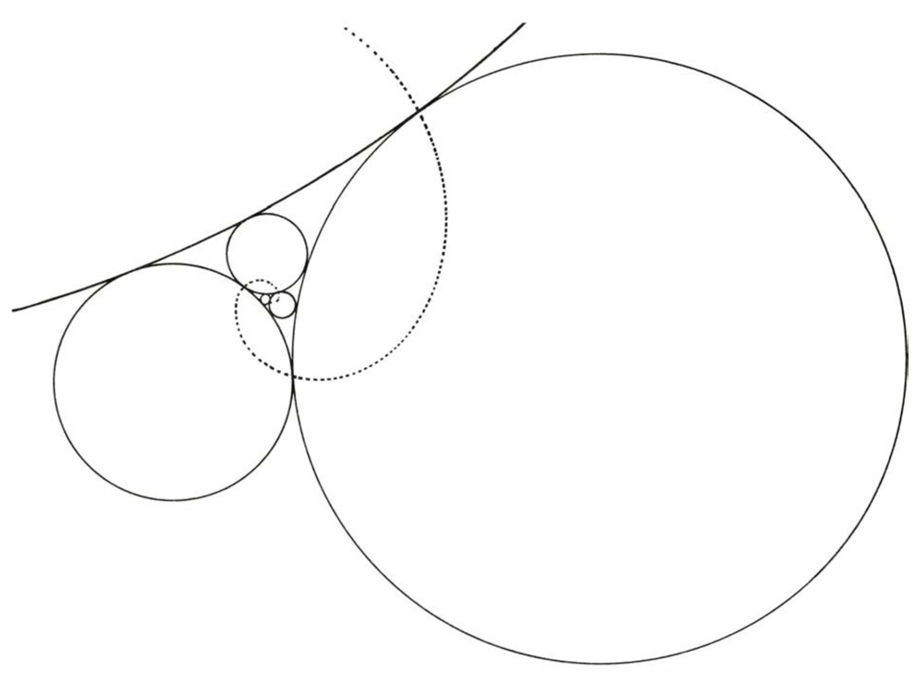
1. *A*1-*C*2	9. *A*2- *C*1
2. *C*2-*A*3	10. *B*1-*C*3
3. *B*4-*C*2	11. *C*3-*A*4
4. *C*2-*A*1	12. *A*3-*C*2
5. *C*1-*A*2	13. *C*4-*A*3
6. *A*2-*B*4	14. *A*3-*B*1
7. *A*4-*C*3	15. *C*2-*A*3
8. *C*3-*A*2	16. *A*3-*C*4

Dije en esta sección el pasado mes de noviembre que no tenía noticia de que se hubiesen comercializado los 30 cubos de MacMahon. La Educational Products Division of International Technical Associates (P.O. Box 7190, Menlo Park, Calif. 94025) me informa de que han fabricado dicha colección, que pueden suministrar por 19,75 dólares más gastos de envío e impuestos (caso de haberlos) a cualquier lugar del mundo. En la sección de diciembre sobre fracciones egipcias di el desarrollo ternario 5/121 = 1/25 + 1/759 + 1/208725. No es posible descomponer esta fracción con menos de tres términos, pero, en cambio, preguntaba si sería posible mejorarlo en el sentido de obtener otro desarrollo ternario que tuviese máximo denominador más pequeño. Sin Hitotumatu, Alan F. Berndt y Pierre Tougne fueron los tres primeros lectores que hallaron la descomposición "óptima": 5/121 = 1/33 + 1/121 + 1/363. Hubo muchísimos más que descubrieron esta solución; demasiados para poder citarlos aquí. C. S. Metchette, Morton Raff y R. Bruce fueron los primeros entre otros muchos que descubrieron el desarrollo óptimo –cuatro términos– para 8/11. Es el siguiente: 8/11 = 1/2 + 1/6 + + 1/22 + 1/66, muy superior al que di en el apartado de soluciones de mi sección del mes de enero. En la cuarta de las respuestas, *b* es igual a 25, y no a 5. Los desarrollos de tal apartado se refieren ambos a 3/25, y no a 3/5, como se escribió en el primero de ellos.

Tengo la satisfacción de informar que un problema largo tiempo pendiente, relativo al tatetí (tres en raya) en casillero tridimensional de 4-por-4-por-4 ha sido finalmente resuelto. Con ayuda de un programa de computador, Oren Patashnik, de los Bell Laboratories, ha conseguido establecer lo que hace ya tiempo se sospechaba, pero nunca se había conseguido demostrar completamente: El primer jugador puede ganar siempre si utiliza la estrategia apropiada. El trabajo de Patashnik, que terminó en 1977 en la Universidad de Yale, fue finalmente comprobado el pasado mes de octubre, por Ken Thompson.



Envasamientos de máxima densidad de uno a nueve círculos en cuadrados



Sucesión áurea de circunferencias de H. S. M. Coxeter

Taller y laboratorio

Microfotografías claras y brillantes con un equipo sencillo y económico

Jearl Walker

El mundo oculto que revela el microscopio ha fascinado a muchos científicos aficionados. Durante los últimos 20 años, varios de estos animosos investigadores han llegado a adquirir una notable habilidad para fotografiar lo que observaban a través del microscopio. Este mes analizaremos algunas técnicas que permiten a los diestros en el uso del microscopio sacar tales imágenes fotográficas.

Podemos encontrar información sobre la microfotografía en diferentes fuentes. La mía procede de James Bell, de Allston, Massachusetts. Ya han aparecido fotografías suyas en *Investigación y Ciencia* y, en esta ocasión, nos ha suministrado diversas muestras de lo que su técnica microfotográfica es capaz de conseguir (tales ejemplares los hemos reproducido en las dos páginas siguientes). La variedad de especímenes que pueden observarse y fotografiarse es casi infinita, de ahí que los ejemplos de Bell los demos sólo a título de introducción. Aunque los fotógrafos profesionales suelen trabajar con equipos muy complejos y caros, las fotografías realizadas por Bell muestran que hasta con un equipo sencillo y económico pueden hacerse fotografías de sorprendente belleza y de gran interés científico.

Bell trabajó con un microscopio monocular Bausch and Lomb fabricado en 1915, equipado con unos viejos objetivos Leitz de 10 y 100 aumentos y unos oculares con un poder de aumento de 10, 7 y 2,5. Disponía de un espejo para dirigir la luz hacia un condensador, el cual hacía que toda la luz convergiera sobre la muestra. Bell quitó el objetivo de su cámara Asahiflex de 35 milímetros, e incorporó la misma al microscopio con un adaptador especial para microscopio, que se encuentra en las tiendas del ramo y que sirve para evitar el paso de la luz del exterior. Enfocó el microscopio de manera que la imagen se formara sobre el cristal del fondo de la cámara. Para evitar los movimientos a la hora de hacer las fotos, Bell accionaba la cámara con un cable disparador. Sacó las fotos con velocidades de obturación entre 1/500 y 1/25 segundos, empleando películas Ektachrome–X o bien Ektakrome–64.

Hay que usar luz muy intensa, pues sólo una pequeña parte de ella llega a la película. Bell se servía para la iluminación de un proyector de diapositivas normal con una lámpara de 400 watt. Para colimar la luz sustituyó el objetivo del proyector por otro que concentraba la luz. En caso de trabajar con menos luz que la empleada por Bell, es necesario incrementar el tiempo de exposición en un segundo o incluso más. Más difícil resulta obtener fotografías claras; con objetos en movimiento, lo más probable es que sólo consigamos un borrón. Si para una fotografía en particular, la luz es demasiado intensa, Bell la atenúa colocando filtros de transparencia uniforme en la trayectoria de los rayos procedentes del proyector. Dichos filtros pueden comprarse en las casas de artículos fotográficos o pueden ser fabricados fotografiando una hoja de papel blanco y revelando el negativo.

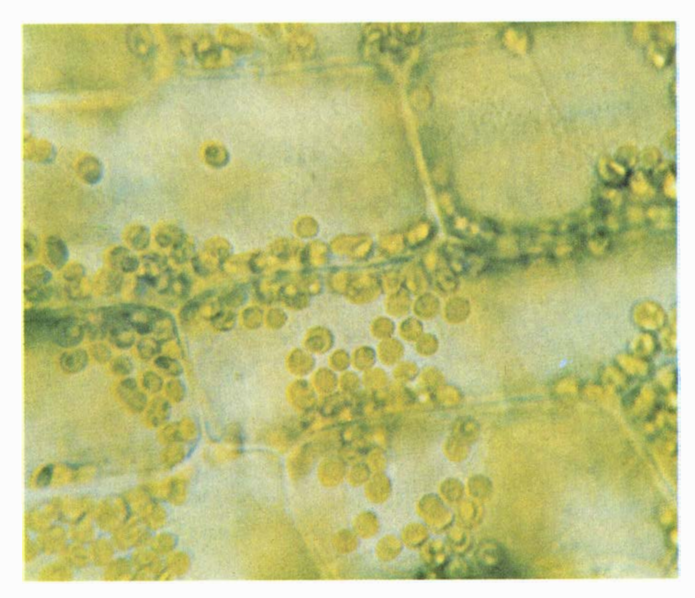
El componente infrarrojo del rayo de luz puede dañar a las muestras, sobre todo si se trata de organismos vivos. Para eliminar dicho componente nocivo de los rayos, puede colocarse un recipiente con agua (de caras paralelas para evitar reflexiones) o bien diversas piezas de cristal óptico en la trayectoria del rayo de luz. El proyector puede incorporar un sistema de refrigeración con idéntica finalidad práctica.

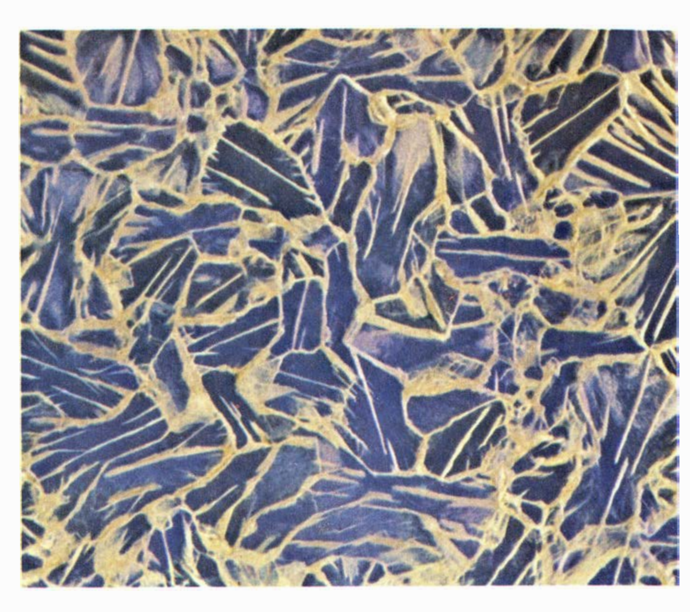
La iluminación con fondo claro es la más común de los diversos tipos de iluminación que se pueden emplear, pero tiene el inconveniente de que destaca poco los detalles de las muestras. Bell logra este tipo de iluminación orientando el espejo de manera que el condensador situado debajo del soporte del espécimen en el microscopio haga que toda la luz converja en dicha muestra. Obrando así, se verá la luz transmitida a través del espécimen. Esta técnica no podrá aplicarse, pues, ni con especímenes gruesos ni opacos, e incluso con muestras delgadas no siempre se distinguirá bien su estructura interna. Por otra parte, los colores naturales de los elementos u organismos a observar, se pierden frecuentemente con el fondo claro. De todos modos, esta técnica puede resultar útil para fotografiar muestras finas en blanco y negro.

Si se quiere obtener la mayor resolución posible, aspecto crucial en trabajos de mucho aumento, deben secundarse las directrices que se indican, respecto a la iluminación Köhler, en el libro *Photography through the Microscope*, auténtica biblia de la microfotografía. Dicho libro ha sido publicado por la Eastman Kodak Company. El manual le dirá cómo ajustar el espejo, el colector y el condensador, de manera que el filamento de la lámpara esté enfocado sobre el condensador y éste a su vez enfoque la imagen del colector sobre el soporte de la muestra. (El colector es la lente situada entre el espejo y la fuente de luz.) Con este método de enfoque la muestra queda uniformemente iluminada y no se ve la imagen enfocada del filamento.

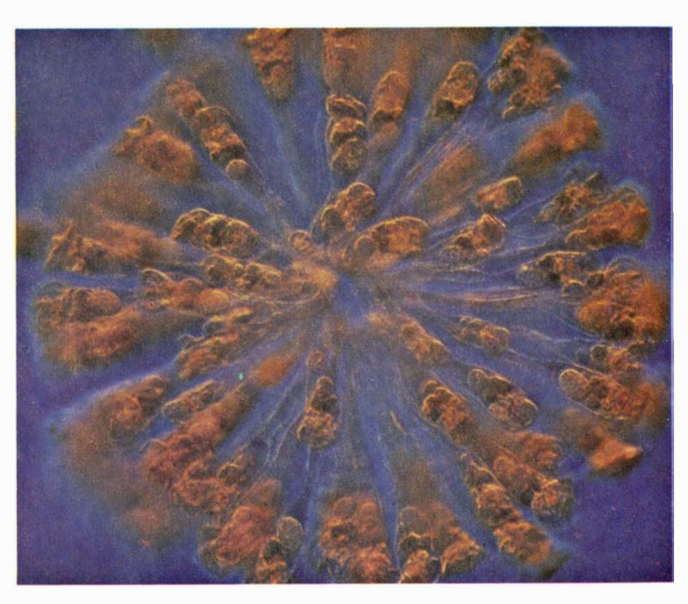
La primera de las fotografías de Bell, que aparece en la ilustración de la página siguiente (arriba, a la izquierda), constituye es un ejemplo de la técnica de fondo claro. Bell empleó 1100 aumentos para fotografiar células vivas del alga de agua dulce *Elodea*, poniendo de manifiesto los distintos cloroplastos. Momentos después de haber tomado la fotografía, los cloroplastos comenzaron a girar dentro de las células conforme iba produciéndose en ellos la fotosíntesis a partir de la luz procedente del proyector. Aunque algunos detalles salen en la fotografía, la imagen no es tan nítida como en otras fotografías, debido en parte a que ese elevado aumento redujo la profundidad de campo.

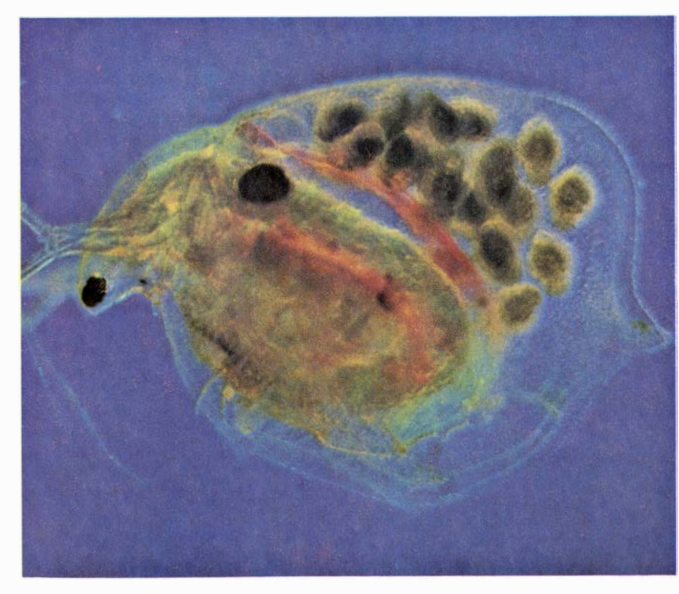
Puede obtenerse un buen contraste en la foto si colocamos la muestra sobre un fondo negro. Para conseguir dicho fondo mientras el objeto a observar seguía estando iluminado, Bell puso un punto negro en un filtro de cristal blanco, que situó justo debajo del condensador instalado bajo el portaobjetos. Tenía bien abierto el diafragma del condensador de suerte que suministrara un ancho cono de luz. El punto negro tenía por misión proyectar una sombra sobre la muestra a fotografiar. La luz que sobresalía por los bordes del punto iluminaba la muestra con luz blanca, por lo que sus colores naturales seguían viéndose. Cuando el portaobjetos está vacío, los rayos de luz que llegan al objetivo

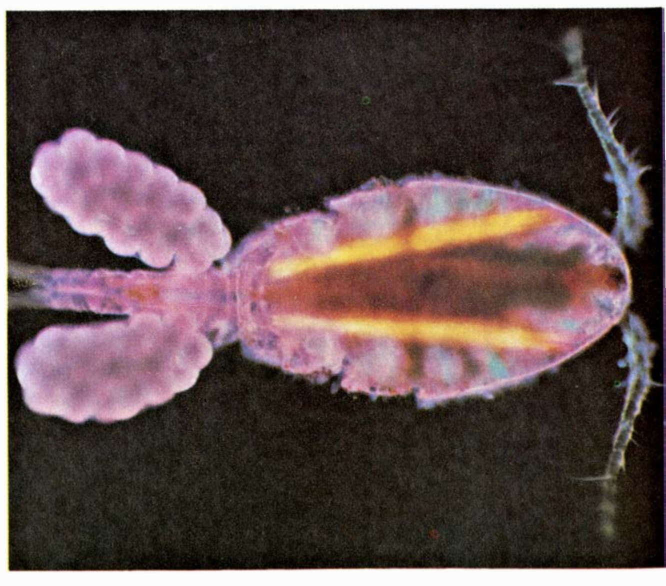

Alga fotografiada con la técnica de campo brillante de James Bell

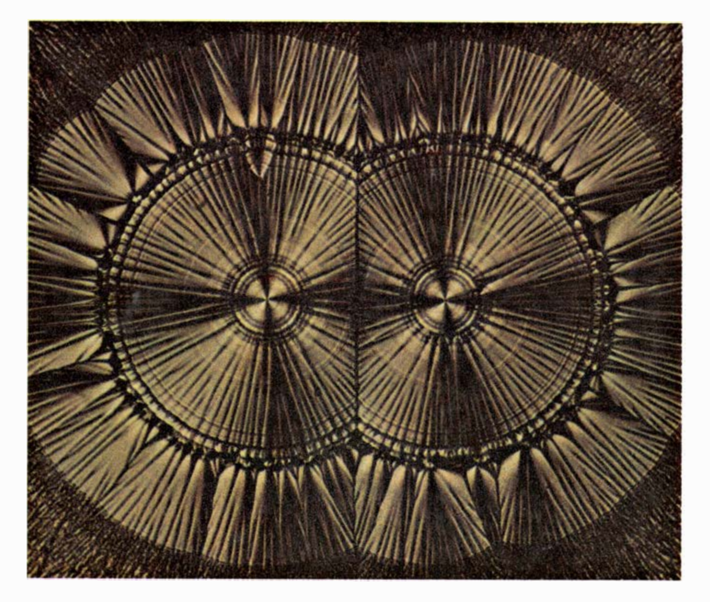

Cristales de éster de colesterol


Embrión de caracol

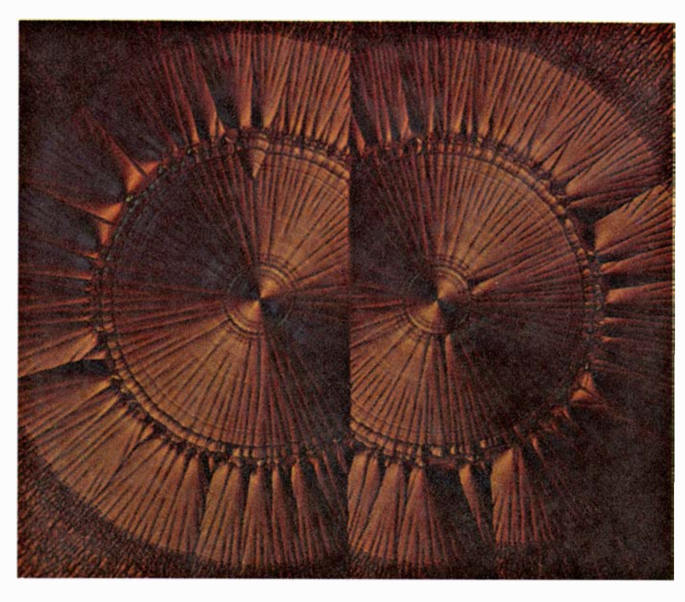

Colonia rotífera de **Conochilus**


La pulga de agua **Simocephalus**


El copépodo **Cyclops**



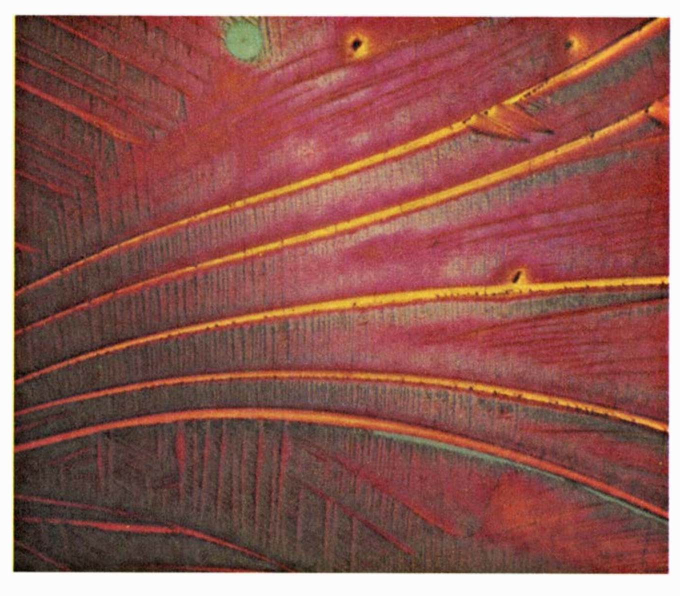
Cristal de vitamina C



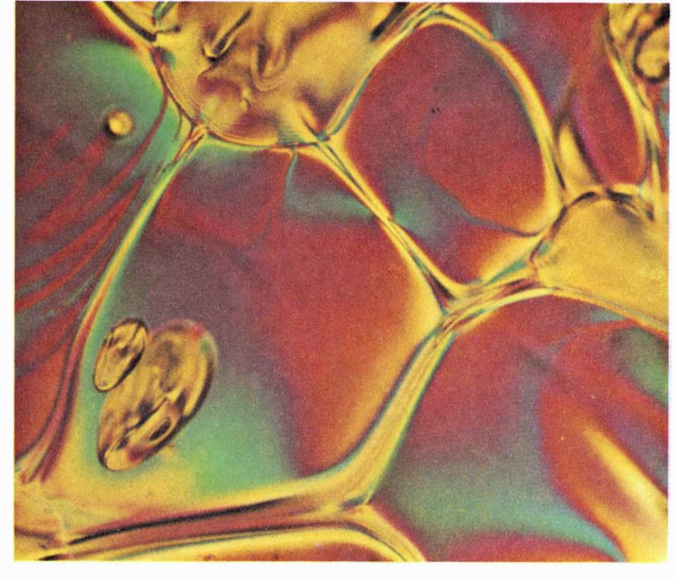
El mismo cristal visto a través de celofán polarizador



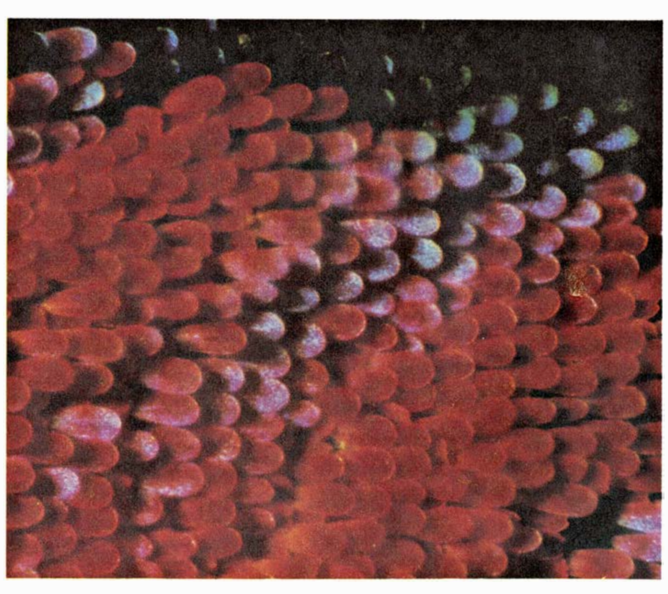
Cristales de ácido hipúrico



Cristales de resorcinol



Cristal líquido nemático



Ala de mariposa

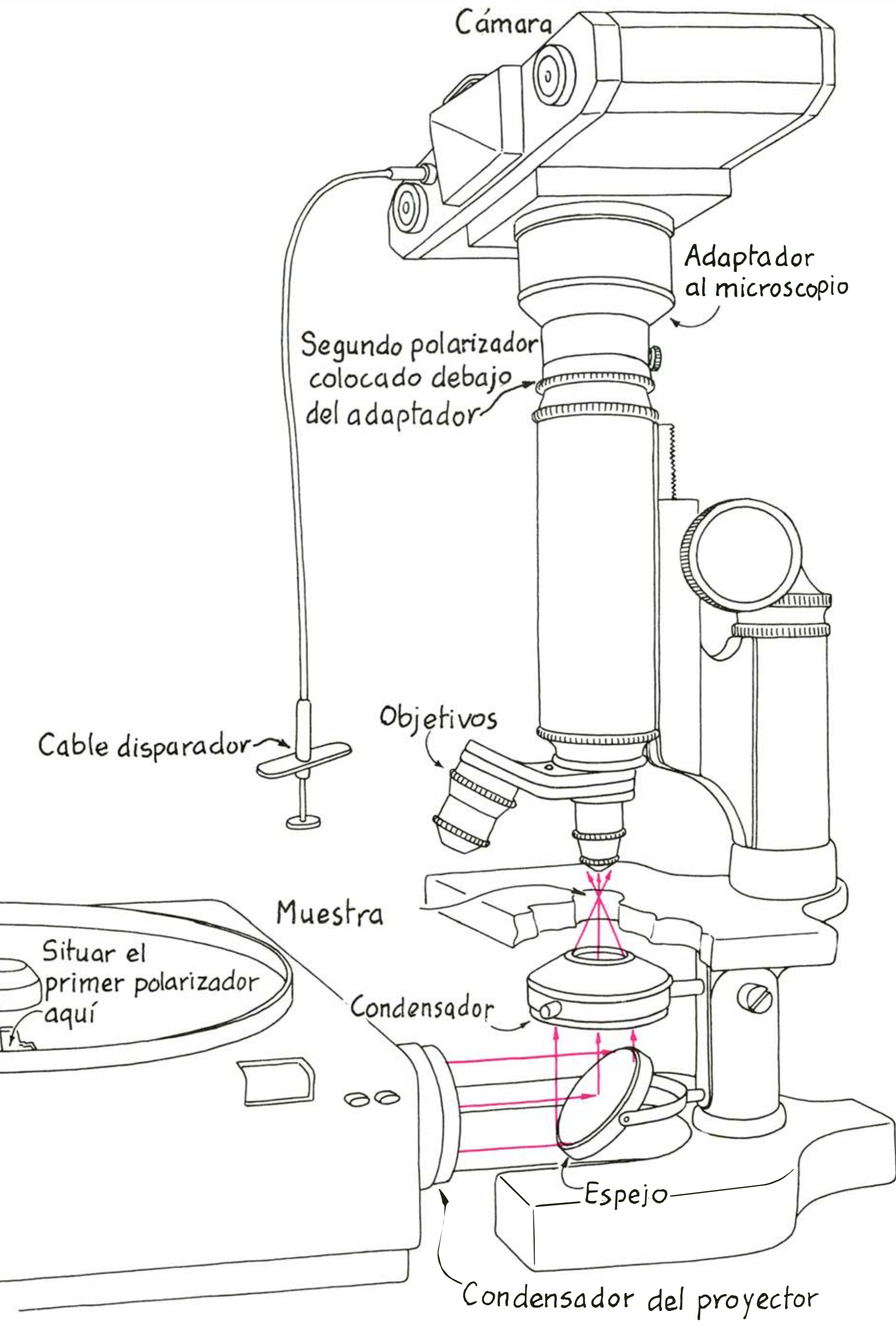
del microscopio han sufrido una divergencia tal que no entra ninguna luz por el tubo; ello determina que ni el observador ni la cámara vean más que un campo obscuro. Con la muestra en el portaobjetos, la luz se dispersa desde él hacia el interior del tubo, permitiendo que el espécimen pueda contemplarse. Esta técnica requiere una intensa fuente luminosa, ya que sólo llegará a la cámara una pequeña parte de la luz inicial.

El filtro para producir el fondo negro se construye pegando un pequeño círculo opaco de cartulina negra sobre un filtro de cristal transparente. El diámetro del círculo debe ser de cinco milímetros o algo más. Conviene determinar el tamaño óptimo probando varios en su propio microscopio. El campo debe verse obscuro cuando no haya nada en el portaobjetos, y aparecerá iluminado cuando coloquemos un espécimen para su observación.

Bell domina la técnica de la iluminación con fondo obscuro con sólo su objetivo de bajo aumento (10 diámetros). Un aumento mayor da lugar a fotografías con un contraste menor debido a la dificultad de mantener la lente que hace de objetivo dentro del cono obscuro proyectado por el punto opaco. Para enfocar correctamente, los objetivos de muchos aumentos deben aproximarse bastante a la muestra; la lente empieza luego a interceptar algunos rayos de luz que provienen de los bordes del punto opaco. Para trabajar según la técnica de Bell con grandes aumentos, habría que comprar un condensador de fondo obscuro, que proporciona un cono de obscuridad más fino.

Con su técnica de fondo obscuro, Bell fotografió un espécimen de *Hydra* y algunas colonias de *Volvox* que había encontrado en estanques locales. Se puede encontrar una relación completa de las muestras útiles para su observación en libros especializados que suelen venderse en los museos de historia natural. (Puede consultarse, con provecho, el *Atlas de microscopia,* de J. Bernis Mateu, 1975.)

Puesto que la técnica fotográfica de Bell puede aplicarse a organismos vivos, si uno tiene suficiente paciencia puede obtener fotografías diferentes de una misma muestra en situaciones diversas, de acuerdo con su comportamiento característico. Así, podemos fotografiar una hidra comiendo. Dándole pequeños gusanos de agua o bien diminutos trozos de carne cruda, se puede conseguir que abra la boca y despliegue sus tentáculos. Headstrom refiere que con una pequeña cantidad de ácido acético o verde de metilo consigue provocar que la hidra descargue uno de sus nematocistos. Estos tubos llenos de púas son lanzados con tal fuerza que pueden penetrar incluso en la presa del animal.



Sistema experimental de Bell

Los *Volvox* forman una colonia esférica gelatinosa, constituida por cientos de organismos flagelados. Si se mira desde muy cerca a una de estas colonias, puede apreciarse que su superficie está como bordada por flagelos, que son los apéndices en forma de latiguillo mediante los cuales los individuos flagelados pueden autopropulsarse. Headstrom recomienda recolectar muestras de este tipo con un tubo de cristal. Una vez que haya localizado una muestra en el agua de un estanque o de una charca, para lo cual quizá sea conveniente ayudarse con una lupa, introduzca el tubo en el agua manteniéndolo tapado con el dedo. Cuando el espécimen esté lo suficientemente cerca de la boca del tubo, quite el dedo; probablemente, la muestra se verá arrastrada hacia el interior del tubo por la corriente de agua que se forma al llenarse. Vuelva a tapar el tubo con el dedo y sáquelo del agua. A continuación, con mucho cuidado, vierta el agua sobre el portaobjetos del microscopio hasta que el espécimen quede depositado sobre el mismo. Para estas operaciones conviene usar portaobjetos especiales que tienen una pequeña concavidad donde queda recogida la muestra con una pequeña cantidad de agua.

Algunas muestras contrastan mejor

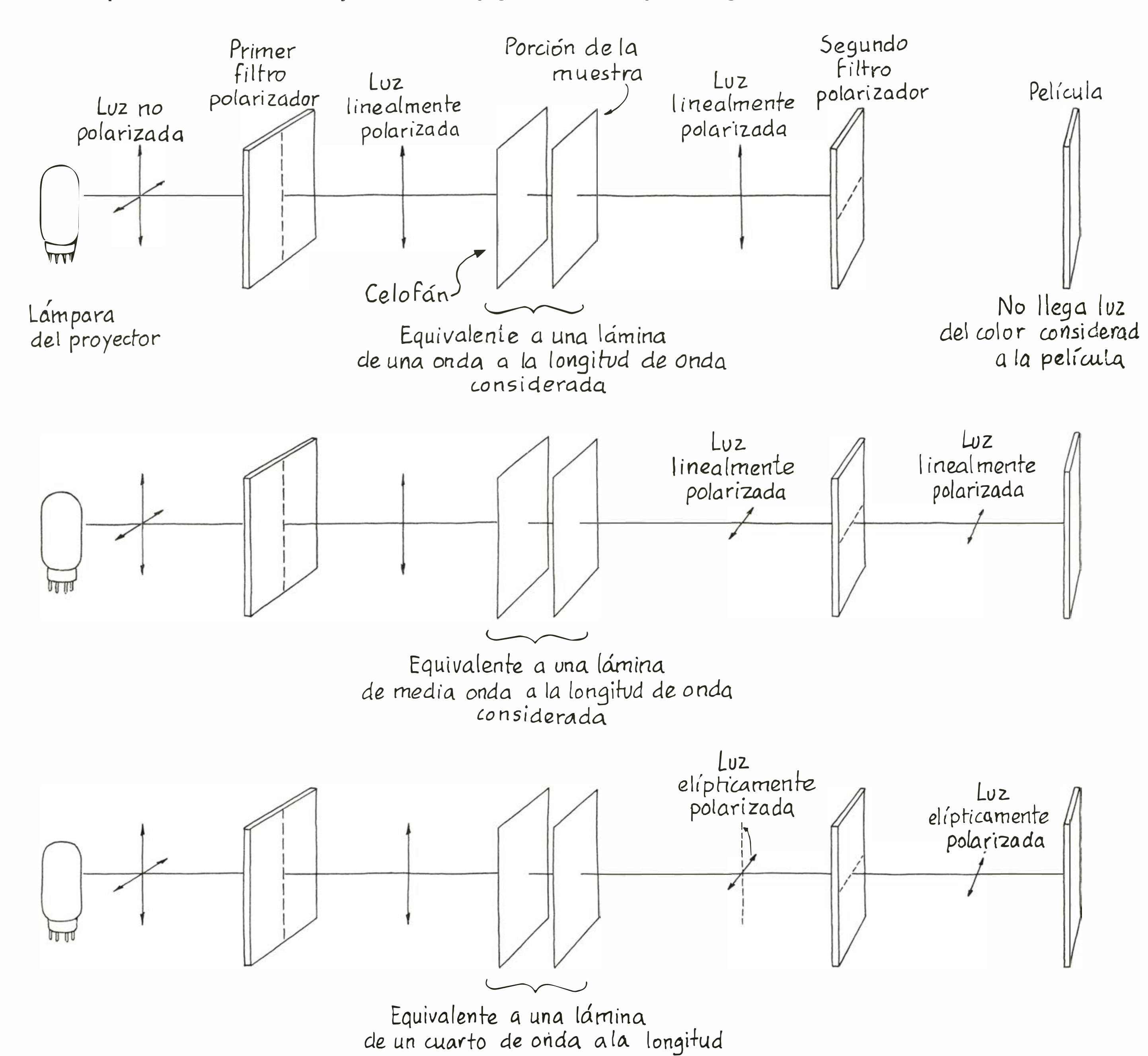
con un fondo de color que con un fondo obscuro. Para conseguir este tipo de colorido (llamado iluminación parcial Rheinberg), Bell sustituyó el punto opaco negro por un punto transparente de color. Por ejemplo, para conseguir un buen contraste con una muestra roja, Bell utilizaba un punto azul o verde, que recortaba de unos filtros de plástico de dichos colores. El espécimen continúa apareciendo con su color natural, ya que es iluminado principalmente por la luz blanca procedente de la zona exterior al punto.

Un fondo azul sirvió a Bell para fotografiar unos cristales de un éster de colesterol que habían recristalizado bajo el cubreobjetos. El fondo permite ver la interfase entre los cristales y las grietas provocadas por las tensiones internas dentro de cada cristal. Para fotografiar embriones del caracol de agua dulce cuando aún se encuentran dentro del huevo, Bell empleó un punto verde con un aumento de 80 diámetros. Estos delicadísimos detalles se hubiesen perdido totalmente con una iluminación con fondo brillante.

Con una iluminación Rheinberg completa, Bell puede seleccionar cualquier color para iluminar la muestra y puede elegir un color diferente para el fondo. El único cambio que debe sufrir el método es pegar un anillo de plástico ligeramente coloreado alrededor del punto más obscuro. La luz que atraviesa el punto da entonces el color del fondo, mientras que la luz procedente del anillo colorea la muestra.

A 120 aumentos, Bell fotografió con un punto azul y un anillo amarillo una colonia de rotíferos *Conochilus*. En una carta que el propio Bell me envió, me decía que los rotíferos habrían sido punto menos que invisibles con una técnica de iluminación de fondo brillante: el contraste de colores resulta fundamental para distinguir sus contornos y su estructura interna.

Bell asegura que la técnica más eficaz para obtener el mejor contraste de colo-



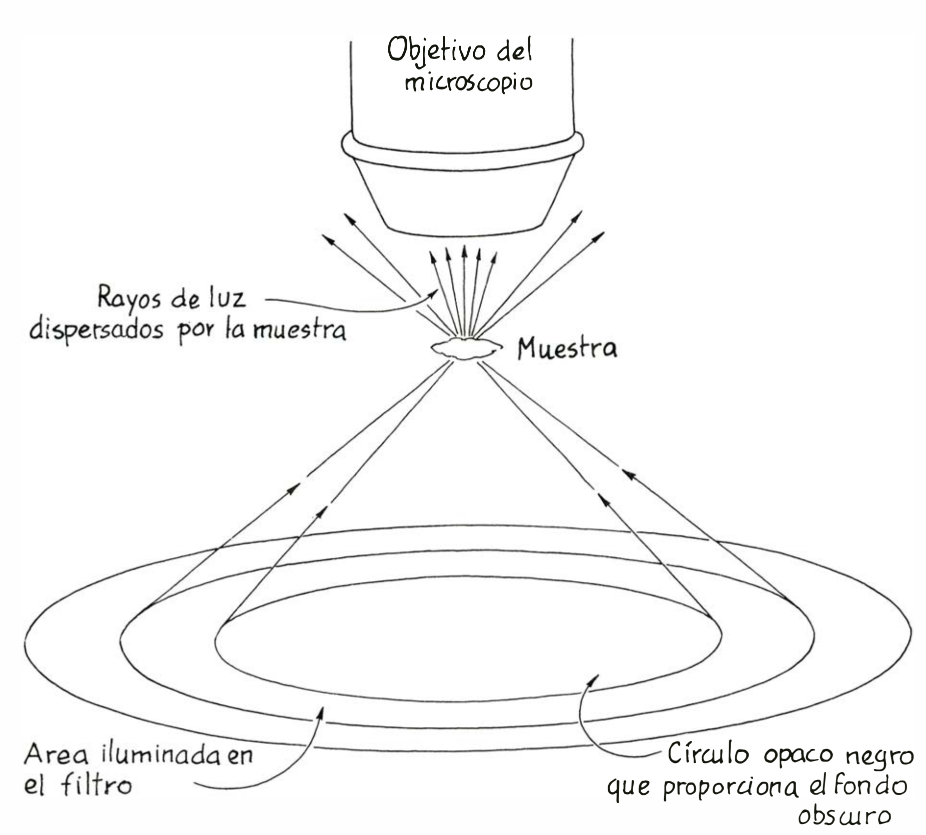
Cambios de la polarización debidos a la birrefringencia

res es hacer las fotografías usando luz polarizada. Esta técnica de trabajo sólo sirve con materiales birrefringentes, tales como los cristales, los músculos y algunos pequeños organismos pluricelulares. Se coloca un filtro polarizador en el proyector de diapositivas (como si fuera una diapositiva normal), con lo que el aparato suministra luz polarizada. Se introduce un segundo filtro en el ocular del microscopio, de manera que se pueda girar hasta que el campo de visión a lo largo del microscopio sea o bien obscuro o brillante, según que el filtro bloquee o deje pasar la dirección de polarización de la luz que viene a través del aparato.

Cuando se introduce un material birrefringente en el portaobjetos, la luz polarizada producida por el primer filtro altera su polarización cuando pasa a través del material. Si la alteración es adecuada, parte de la luz puede pasar entonces a través del segundo filtro polarizador, en el ocular, aun cuando el filtro bloquearía la luz. (En esta misma sección, en febrero de 1978, se trataron los pormenores de cómo un material birrefringente cambia el sentido de polarización.) En resumen, la luz se transmite a través de este material cuando parte de la misma está polarizada a lo largo de un eje y la otra polarizada a lo largo de un eje perpendicular. Ambos ejes son perpendiculares al rayo de luz. La velocidad de la luz es mayor en uno de los sentidos de polarización que en el otro. En razón de esta diferencia, los dos sentidos de polarización pueden salir procedentes del material fuera de fase. La polarización emergente está determinada por la combinación de dos sentidos de polarización: en consecuencia, la luz que sale podría estar polarizada de varias formas.

Si los dos sentidos de polarización salen con una diferencia de fase de media longitud de onda, la luz que emerge está otra vez linealmente polarizada, pero ahora lo estará a lo largo de un eje perpendicular a la dirección de polarización de la luz incidente. Esta luz que sale puede ser transmitida o bloqueada por el filtro polarizador situado en el ocular, según la orientación de éste.

Si la birrefringencia ha obligado a las direcciones de polarización que salen a estar desfasadas un cuarto de longitud de onda, la luz que emerge estará polarizada elípticamente; ello significa que la punta del vector de polarización gira alrededor del rayo de luz, describiendo una elipse. Esta luz pasará por el filtro situado en el ocular, con independencia de la orientación de éste. Si los sentidos de polarización emergentes coinciden, la luz transmitida estará polarizada de la misma forma que la luz incidente. La orientación del segundo filtro determinará entonces si se ve o no luz alguna.



Preparación de un fondo obscuro

El que se obtenga un resultado u otro depende del espesor y birrefringencia de la muestra y de la longitud de onda de la luz. Para una región concreta de la muestra, el extremo rojo del espectro de la luz blanca incidente sobre él puede acabar siendo bloqueado por el segundo filtro polarizador, mientras que puede transmitirse el azul. Esta zona aparecerá de color azul para el observador. Puede aparecer otra zona de color amarillo cuando estén bloqueados los otros colores de la luz blanca incidente. La ventaja de usar así luz polarizada estriba en que la variación de color que se produce en la muestra, que de otra forma sería incolora, permite al observador distinguir sus perfiles y su estructura.

Bell me ha enviado dos fotografías hechas con luz polarizada. En la primera, tomada a 120 aumentos, se aprecian cristales de ácido hipúrico que se habían disuelto en isopropanol y se recristalizaron luego en la diapositiva. La muestra era lo suficientemente birrefringente y no uniforme como para producir muchos colores. En el otro ejemplo (a 150 aumentos) se había disuelto cristales de resorcinol en acetona, que se quemó después para facilitar la recristalización rápida.

Se pueden producir más contrastes de color poniendo uno o más trozos de papel de celofán o de cinta adhesiva transparente en el soporte del filtro por debajo de la muestra. Ambos son materiales birrefringentes que alteran la polarización de la luz. En lugar de celofán podemos usar un trozo de mica fina que puede sacarse fácilmente de un fragmento mayor con una cuchilla. Se supone que cualquiera de estos materiales actúa como filtro retardador, porque obliga a que un sentido de polarización quede detrás del sentido perpendicular. Un filtro de retardo es útil para aumentar la variación de color en un modelo débilmente birrefringente. No todos los tipos de celofán habrán de valer, por lo que será preciso probar varios hasta dar con el adecuado. También podrían usarse hojas de plástico del tipo de las que se emplean para envolver alimentos, siguiendo las normas que expuse en mi artículo sobre birrefringencia.

Bell ha presentado algunos otros ejemplos de cómo usa la luz polarizada. Hizo una fotografía de 100 aumentos a una pulga de agua viva con su cría poniendo un punto azul de fondo y con celofán y filtros polarizadores para destacar la estructura interna del animal. La mancha del ojo queda en negro, y el tubo digestivo en naranja.

A 50 aumentos, fotografió de una forma similar el copépodo *Cyclops*, con

la diferencia que usó un punto negro para proporcionar un fondo obscuro. El animal aparece con la mancha de los ojos en rojo, dos bandas musculares en amarillo y el tubo digestivo en marrón. Los dos sacos laterales de forma de pera contienen huevos. Muchos de estos detalles se hubieran perdido en un campo iluminado normalmente.

Puede comprobarse el efecto del celofán en dos fotografías de Bell, ambas de ácido ascórbico (vitamina C) recristalizado por él. Las fotografías muestran los dos mismos cristales adyacentes, y se tomaron exactamente igual, con la salvedad de que se utilizó celofán para la segunda fotografía. Adviértase que en la última el color adicional se debe a la birrefringencia del celofán.

Bell preparó los cristales disolviendo ácido ascórbico en isopropanol, puso luego un poco de esta solución en un cristal portaobjetos y quemó después el alcohol. Los cristales se formaron a los pocos segundos. Una vez sacada una serie de fotografías, puede volverse a disolver el ácido ascórbico y empezar de nuevo. Esta técnica produce cristales lo suficientemente finos como para que resulten transparentes; por tanto, pueden fotografiarse fácilmente. Otro método de obtención de cristales es dejar que el disolvente se evapore lentamente.

El lector puede conseguir también una cristalización fundiendo la sustancia en un portaobjetos sobre una llama. Aquélla cristalizará al enfriarse. Bell no es muy partidario de esta técnica, porque el punto de fusión de algunas sustancias es demasiado alto, y ello hace que se rompan muchos portaobjetos. Sin embargo, señala que, si se quieren estudiar ciertos tipos de cristales líquidos, habrá que recurrir a dicha técnica, enfriando la sustancia lentamente desde su punto de fusión para conseguir el estado de líquido cristalino. La fotografía de Bell de un cristal líquido nemático se sacó a 100 aumentos, a temperatura ambiente, tras haber disuelto el material en acetona. El movimiento en esta recristalización fue tan rápido que tuvo que disparar la fotografía a 1/500 segundos para retener la actividad.

Debido a su espesor algunas muestras no pueden fotografiarse con luz transmitida; el obstáculo se salva con luz reflejada. Para ello, Bell apoya el proyector en unos libros y ajusta el ángulo de la luz incidente hasta que obtiene la iluminación adecuada en la muestra. Un ejemplo del resultado es su fotografía del ala de una mariposa. Muchos de los colores que se ven en las alas de las mariposas, en los élitros de los escarabajos y en las plumas de las aves no se deben a la pigmentación, sino a la interferencia de las ondas de luz. Algunas alas de mariposa están formadas por capas de cutícula transparente. Cuando se refleja la luz de una de las capas, parte de ella procede de la reflexión en su superficie externa y otra parte proviene de su superficie posterior, después de haber atravesado la capa en cuestión. Los dos rayos emergentes pueden interferir entre sí para producir colores, como ocurre en las pompas de jabón (he descrito estas interferencias en esta misma sección en el número de noviembre). Quizás el observador quiera ver varios tipos distintos de alas de mariposa. Estas pueden comprarse en casas dedicadas a muestras empleadas en prácticas de ciencias naturales y en tiendas de objetos de regalo en muchas ciudades.

La interferencia de colores resultaría ser muy débil si el ala estuviera iluminada sólo con la luz transmitida. Los colores tendrían que aparecer, entonces, como consecuencia de la interferencia entre la luz transmitida por la capa similar a la cutícula (sin reflexión interna) y la luz doblemente reflejada en el interior de la capa, antes de salir. Esta luz será tan débil, con relación a la luz no reflejada, que producirá poca interferencia (y, por tanto, poco color).

Para encontrar la exposición ideal hay que controlar los disparos con una gama de tiempos de exposición o bien incorporar un medidor de luz en el microscopio. Bell usa un medidor de luz casero compuesto por un voltímetro digital capaz de detectar hasta varios milivolt y una fotocélula de silicio. Bell montó la célula en una carcasa de cartón que se encajaba sobre el punto de mira de la cámara. A medida que cambia la intensidad de luz con la muestra y el modo de iluminación, la resistencia de la fotocélula varía inversamente. A través de los circuitos intermedios, estas variaciones modifican la lectura en el contador de milivolt. Tras varios experimentos, Bell consiguió construir una tabla de lecturas de manera que fuera posible interpretarlas inmediatamente en función del tiempo de exposición. (En el número de junio de 1977 de la revista *Popular Electronics* apareció diseñado un medidor de luz simple, aunque muy sensible, que podía construirse sin gran complicación con un fotorresistor de sulfuro de cadmio.)

Si se quiere obtener una buena reproducción de los colores naturales de un modelo, conviene atinar en la elección del tipo de película de color que se use; debe estar ajustada a la distribución de colores de la lámpara. Todas las lámparas parecen blancas para el ojo humano, pero la distribución real de intensidades a través del espectro visible depende de la temperatura de la superficie emisora de la lámpara. Una lámpara de luz blanca, de baja temperatura, tiene menos bajas frecuencias de luz visible (es decir, el final azul del espectro) que una lámpara de temperatura más alta. Las distintas películas de color se han ajustado de alguna forma para compensar esta diferencia de distribución del color de la luz blanca. Sin embargo, puede haber alguna discrepancia entre los colores reales de un modelo y lo que se ve en una fotografía en color. Se puede remediar la situación poniendo un filtro de color delante del proyector (o cualquier otra fuente de color). El libro de Kodak explica qué filtros deben emplearse.

Bell tiene una lámpara DAT de 400 watt en su proyector. La temperatura de la superficie del filamento emisor es tan alta que la luz semeja la luz solar en cuanto a la distribución de color. Por ello Bell utiliza película de color especial para la luz solar en vez de las películas que tienen un equilibrio de colores calculado para lámparas de tungsteno que funcionen a temperaturas superficiales inferiores.

Teresa Owens, antes de su graduación en el Reed College de Portland, Oregon, hizo un trabajo muy detallado y exhaustivo sobre el índice de enfriamiento del agua partiendo de varias temperaturas iniciales. Como ya dije en esta misma sección en noviembre de 1977, el agua caliente alcanza a veces el punto de congelación antes de lo que lo haga el agua inicialmente fría. El experimento es tan rico en variables que resulta difícil sustraerse a la tentación de averiguar sus causas.

Con un equipo experimental similar al mío, Owens registró la temperatura del congelador y la de varios puntos de un recipiente de agua depositado dentro del mismo. Sus resultados indican que el efecto se da probablemente, aunque puede perderse sin mayores dificultades con la variación de la temperatura del ambiente entre las distintas pruebas del experimento. De todas formas, de los datos obtenidos saca la conclusión de que los dos factores más importantes que determinan el efecto antedicho son: la temperatura inicial del ambiente y la circulación del aire sobre la parte superior del recipiente. En su opinión, el gradiente de temperatura dentro del agua y la masa perdida por evaporación no importan demasiado. Queda mucho por aprender sobre este experimento. Si algún lector aficionado llegara más allá en sus ensayos, me gustaría conocer sus conclusiones.

Libros

Medicina cuantitativa, psicología como ciencia, energía solar para edificios y citología básica

Philip Morrison, Luis Alonso, Miguel Fisac y Alfredo Carrato

NON-INVASIVE CLINICAL MEASUREMENT, edición a cargo de David Taylor y Joan Whamond, University Park Press. Los murciélagos y los cetáceos antes que nosotros, pero ahora también los humanos, usan rutinariamente impulsos de ultrasonido (o de microondas) para cartografiar la noche o las profundidades. Alrededor de la mitad de los ensayos de este simposio de expertos deja al lector en la frontera de los avances recientes en la exploración, mediante el sonar, de ese mar entrañable e interno que es el propio cuerpo humano. Todo médico que se tome en serio su compromiso jurado de al menos no dañar, debe cuestionar la lesión producida secundariamente por muchas clases de sondeos que son valiosos para el cuidado racional de los enfermos. Las puntas de agujas, los tubitos agresores y el haz ionizante de los rayos Roentgen pueden comportar riesgos y dolor para el paciente al que se pretende ayudar con la información que proporcionan. De ahí los instrumentos "no agresivos" descritos en esta colección actualizada de 15 estudios detallados.

A los impulsos ultrasónicos de unos cuantos megahertz se les ha hallado empleo, desde hace algunos años, para cartografiar el cuerpo vivo desde fuera. El dispositivo explorador en modalidad *A* de la jerga bélica para el radar, en la cual la huella del osciloscopio traza el tiempo de demora de retorno del eco en función del azimut, fue quizás el primero. Hoy en día, la modalidad *B*, con un haz ultrasónico explorador cuya intensidad de eco se señala punto a punto para construir una imagen bidimensional brillante en la pantalla, tiene ya diez años de edad y se usa mucho. Por ejemplo, el cráneo del feto puede medirse en el útero y predecirse la fecha del parto con una desviación típica de unos 12 días.

Los autores proceden de grupos de bioingeniería, dispersos entre Toronto y Bucarest. La formación de la imagen es sólo parte del asunto. Las penetrantes longitudes de onda sónica empleadas miden apenas unas decenas de veces más que el diámetro de los eritrocitos; la débil dispersión de estas partículas arrastradas por lo corriente no es difícil de detectar, y el corrimiento Doppler de unos cuantos kilociclos se mide fácilmente en el eco. La geometría del flujo es compleja; no es fácil enfocar en derechura el eje de la aorta escondida. Una hábil geometría del dispositivo explorador, con sondas gemelas en ángulo recto, puede proporcionar un corrimiento de frecuencia independiente del ángulo. La superposición de una imagen de exploración *B* a un Doppler de exploración *B* se muestra aún más prometedora. Exige a esas notables pastillas microelectrónicas la manipulación de un montón de datos digitales.

El perfil del flujo en el interior de vasos sanguíneos semiobturados y la diaria medición absoluta de la velocidad de flujo y de la superficie real no están lejos. La ambigüedad del alcance en sistemas pulsantes Doppler ordinarios, producida por el conflicto entre la velocidad de repetición de impulsos y el corrimiento Doppler, limita el producto medible del alcance por la velocidad de la sangre. Para eludirla se está creando sistemas de modulación de frecuencia como el de los murciélagos e incluso modulación de códigos de impulsos, con una elaboración compleja, más usual en el contexto de la guerra antisubmarina que en el pabellón hospitalario.

La cantidad de oxígeno transportada en el torrente arterial se hace marginal durante crisis tales como la anestesia y el nacimiento prematuro. Unas muestras tomadas cada cierto tiempo apenas si pueden cartografiar los cambios previstos; en un individuo normal, la tos puede inducir cambios de decenas por ciento durante uno o dos minutos. En un sistema descrito en la obra se ofrece el continuo registro no agresivo de la concentración de oxígeno en la sangre. Es un subproducto reconocible de la observación microscópica que la mayoría de los que fuimos estudiantes podemos recordar: induciendo espasmos de eritrocitos a lo largo de los capilares de la pata palmada de una rana. Los gases sanguíneos de la circulación arterial profunda se representan con gran exactitud en sangre inducida a correr por capilares cercanos a la piel mediante aplicación local de calor. A través de la piel intacta se difunde algo de oxígeno. Una elegante sonda epidérmica de contacto, con su propio calefactor pequeñito cuidadosamente regulado por termistores, muestrea el oxígeno en fuga mediante una delgada película de teflón y celofán. Dentro de la sonda una diminuta pila de plata, cloruro de plata y platino mide polarográficamente la tensión de oxígeno, que se relaciona linealmente con la corriente reductora de unos cuantos nanoamperios. La sonda reacciona en unos 10 segundos. Es fácil de calibrar, y se reseña una larga serie de correlaciones clínicas. El aparato creado por el grupo de la Universidad de Marburg aquí citado ya se está fabricando.

Queda mucho por recorrer antes de que esta pericia produzca dispositivos fiables y baratos, adecuados para enfermos gravísimos encamados. El director de la edición, miembro del Royal College of Surgeons, ofrece una breve perspectiva del pasado y del futuro. Los diversos ensayos son técnicos y el alcance de cada uno de ellos es limitado. Necesitamos un método mejor para obtener una imagen del feto durante el primer trimestre "cuando debería diagnosticarse la anormalidad". El ultrasonido es demasiado basto. Posiblemente podría usarse un radar de ondas muy cortas, pero la garantía contra efectos teratogénicos debe estar bien fundada para justificar una exposición a microondas en este período de máxima sensibilidad (no se comenta aquí respecto de la seguridad del ultrasonido).

La medicina preventiva exige mayor atención, especialmente para los males de la población cuya edad sea de menos de 40 años. Para avanzar en esto se precisa contar con dispositivos móviles no agresivos de detección a largo plazo, con almacenamiento de datos en cinta magnetofónica o quizá con enlaces de radiotransmisión. La presión, flujo y gases sanguíneos son los principales datos que, a nuestro entender, necesitamos ahora; las señales eléctricas procedentes del corazón y del cerebro ya se están registrando en pacientes ambulatorios. Hay en esto un desafío que debería dar un gran paso en el sentido de ofrecer a

la comunidad de expertos en sistemas e informática el equivalente moral de una guerra. (Ph. M.)

UNA INTRODUCCION AL METODO CIENTIFICO EN PSICOLOGIA, por Ramón Bayés, Editorial Fontanella, Barcelona, 1978. El concepto de ciencia y de sus métodos específicos se remontan al nacimiento del saber humano en la Jonia del siglo VII antes de Cristo. Parménides esbozó los dos caminos que llevaban al ser de las cosas, al ser último, el de la verdad y el de la opinión. Aristóteles regula en los *Segundos analíticos* las propiedades que caracterizan el verdadero saber científico. Cuando la Nuova Scienza y la filosofía racionalista afilen sus armas contra el aristotelismo lo harán, preferentemente, contra el método silogístico, no contra su intento de develar la realidad de las cosas, pretensión que guía evidentemente a todos los pensadores en todos los tiempos.

Problema inmediato al de la identificación de las notas que definen la ciencia y su método estricto es el de la clasificación y unidad o multiplicidad de las distintas ramas del saber que adecuan sus contenidos y sus pautas de investigación al concepto general de ciencia. Hasta la modernidad más avanzada, las ciencias se incardinaban en el tronco común de una supraciencia, de un saber más general, que constituía la filosofía. La emancipación y la independencia en el concierto de las distintas disciplinas fue obligando a nuevas definiciones del concepto de ciencia. Hasta que, por influjo sobre todo del positivismo comtiano, las verdaderas ciencias se consideraron las ciencias físicas, entendiendo por tales aquellas que podían someterse a cuantificación y comprobación experimental o empírica de sus asertos. Y puesto que sólo pueden medirse los fenómenos, lo que aparece, ciencia será el conjunto de conocimientos que averigua dónde, cómo, cuándo y por qué o cómo se presentan los fenómenos. Tiene, pues, la virtualidad de repetir, o al menos simular, los fenomenos que estudia.

La formalización de las ciencias ha traído consigo la posibilidad de incluir entre ellas a la lógica, aunque es verdad que ésta no puede someter a verificación por vía de hecho sus constructos mentales, pero sí puede hacerlo mediante unas reglas o pruebas de coherencia.

Queda, sin embargo, un grupo de ciencias, las llamadas ciencias humanas, que por carecer quizá de un lenguaje único y unívoco, no acaban de entrar en esa categoría. La sociología, la psicología y la historia pertenecen, entre otras, al grupo de disciplinas que piden un sitio en el mundo de la ciencia. El reciente premio Nobel de economía (véase INVESTIGACION Y CIENCIA, enero) ha hecho bastante en este terreno, por la sociología. Ramón Bayés presenta en esta obra un loable esfuerzo por la psicología.

Las limitaciones de la ciencia psicológica por adquirir el rango de ciencia empírica que todo profesor de psicología experimental quiere darle (y Bayés profesa psicología experimental en la Universidad Autónoma de Barcelona) proceden de lo que tradicionalmente se ha venido asignando a la parte mental del hombre y, en un grado mucho más elevado de cuanto tiende a pensarse, de su base neurológica. El cerebro humano es una incógnita en muchos aspectos. Los avances registrados en las máquinas automáticas o computadores resultan insignificantes ante las posibilidades de nuestras neuronas. El experimento científico puede provocar, con fines demostrativos, la repetición de un fenómeno que siempre debe darse en unos supuestos, de temperatura y presión por ejemplo, determinados. Jamás un acto humano voluntario o consciente es exactamente igual a otro. Las desviaciones de un experimento atribuibles a la calidad de los materiales, o al estado de los elementos que intervienen, pueden corregirse matemáticamente, o al menos preverse con exactitud probabilística. En un acto de pensar, por contra, nunca dos individuos podrán tener *in mente* el mismo campo de inteligencia, ni siquiera uno mismo en dos etapas sucesivas, por inmediatas que sean.

No se trata, pues, de caer en la trampa del introspeccionismo, como opuesto a las metodologías operacionales, que introducen variables sujetas a comprobación y experimentación. El camino hacia la formalización, hacia la categoría de ciencia, debe pasar por estos esfuerzos cuantificadores, sobre todo echando mano a la bioestadística. No me parece que en psicología pueda darse nada de un modo absoluto, sino siempre aproximativo con muchos grados de libertad.

La psicología, por lo demás, no se encuentra sola en esta búsqueda de sus señas de identidad. La década de 1960 fue fecunda en polémicas sobre la reducción de la biología a la fisicoquímica, polémicas que, hoy por hoy, nadie da por terminadas, aunque los filósofos de la ciencia hayan apuntado sus armas hacia otros problemas. Es más, ni siquiera la química se siente segura y son muchos los que la consideran una proyección de la física. La ventaja de estas ciencias empíricas es que tienen acotado su lenguaje, que responde a unos fenómenos únicos e identificables por todos. El libro de Bayés aporta un interesante apéndice para hacerse una idea exacta del panorama actual de la psicología en España. Muy triste por cierto. (L. A.).

ENERGIA SOLAR Y EDIFICACION, por SV Szokolay, Editorial Blume, Barcelona, 1978. El encarecimiento de los crudos petrolíferos ha producido una reacción en cadena que ha afectado a todos los planteamientos energéticos relacionados con la industria [véase en este mismo número de INVESTIGACION Y CIENCIA el artículo de Edward D. Griffith y Alan W. Clarke sobre "La producción mundial de carbón", como una alternativa viable a escala global]. Y al analizar las posibilidades de conseguir energía lo más abundante y barata posible, tenía forzosamente que aparecer un creciente interés por la energía solar.
solar para ciertos usos, como pueden ser los necesarios para sustituir la gasolina en los vehículos automóviles, se aprecia a primera vista que no será fácil y, por eliminación de sustituciones más problemáticas, el estudio del uso de la energía solar se ha centrado, en los últimos años, en soluciones que pueden ser alternativa factible, y hasta económicamente competitiva, en la edificación y sobre todo en sus dos servicios más frecuentes de agua caliente y de acondicionamiento térmico.

Los estudios e investigaciones sobre energía solar para usos relacionados con la edificación han interesado lógicamente a los arquitectos y a todo el sector dedicado a la construcción. Como consecuencia de este interés, han comenzado a proliferar libros y ensayos sobre este tema. El libro que comentamos es uno de ellos, traducido de una segunda edición inglesa, que tiene un carácter de divulgación y recopilación de diferentes ensayos como se indica en el prólogo. Este tipo de publicaciones es muy adecuado para los arquitectos que desean adquirir un conocimiento lo suficientemente elevado del tema como para que les pueda servir para tener criterios claros a la hora de juzgar, analizar y hasta decidirse a realizar ciertos ensayos e investigaciones propias, sin tener que ser especialistas.

En esta línea, los primeros capítulos, los referidos a "Principios Generales", "Métodos de captación", "Usos" y "Sol y Construcción" cumplen su cometido, aunque algunos apartados lleguen a aclaraciones excesivamente elementales.

Pero la esencia del libro es en realidad el capítulo quinto dedicado al estudio monográfico de 38 ejemplos. Y es aquí donde puede apreciarse que es un libro obsoleto, ya que, entre los ejemplos que se analizan, algunos de ellos pertenecen a lo que podríamos llamar la prehistoria de la investigación de la energía solar en los años 1939, 1945, 1947, 1948, 1955, 1956, 1958... y los más recientes de 1972, 1973, 1974 y un proyecto, al parecer sin ejecutar, de 1975.

Desde estas últimas fechas, precisamente, es cuando la investigación en este campo de la energía solar aplicada a la construcción adquiere una mayor aceleración y el problema no queda reducido a un simple diletantismo, y firmas industriales importantes se lanzan seriamente a la investigación de la energía solar para usos arquitectónicos. Resulta poco explicable que se traduzcan al castellano libros de investigación técnica, que por la rapidez con que se sucede la investigación en este campo, al cabo de uno o dos años, ya quedan anticuados, y no se publiquen las más recientes investigaciones realizadas no sólo en el extranjero sino también en España en donde disponemos de ejemplos de 1977 y 1978 de mucho mayor, interés y actualidad. En los capítulos siguientes se mezclan algunos datos interesantes, para el conocimiento general del problema, con referencias algunas de ellas totalmente pasadas o pormenorizadas en latitudes y circunstancias totalmente distintas de las nuestras (aunque no faltan ejemplos hispanoaméricanos).

La falta de información sobre nuestra situación en esta técnica se patentiza al enumerar situaciones de la investigación sobre energía solar en países como la India, y otros. El autor desconoce que el ejemplo más completo e importante realizado hasta hoy en Europa, de edificio de gran ciudad con total y exclusivo acondicionamiento térmico con energía solar por aire, se ha realizado en Madrid hace ya un año. Y otras soluciones de gran envergadura están proyectadas para Madrid y Barcelona. (M.F.)

MANUAL DE CITOLOGIA, por M. Maillet, Toray-Masson S.A. Barcelona, 1978. El autor de esta obra, profesor Maillet, es una de las figuras más relevantes en el campo citológico e histoquímico internacional. No es tarea menor que una citología, que constituye un programa completo respecto a la célula animal, contemple toda proyección posible al campo de la medicina, ofrezca cualquiera de sus tópicos de acuerdo con el estado actual del avance multidisciplinario de la biología celular y sea lo bastante concisa para no abrumar al alumno. No olvidemos que la citología, con toda su indudable importancia, es sólo una pequeña parte dentro de un plan de estudios bien equilibrado en la carrera de medicina. Pues bien, la difícil integración de todos estos factores ha sido llevada a feliz término por el profesor Maillet en esta obra, bien planificada, en lenguaje conciso e ideas claras, completamente actuales y sintonizadas con todos los avances colaterales de la biofísica, bioquímica, citoquímica estructural, citogenética y patología celular. Es un libro original en el que aparece una unidad coherente de materias tratadas con un estilo inmejorable, didáctico e informativo, sin una sola línea innecesaria, sin un solo punto que falte en el contexto global. Las figuras y esquemas son fiel exponente de la claridad de ideas ya mencionada. El glosario final, una ayuda apreciable para el lector novel. (A. C.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA PRODUCCION MUNDIAL DE CARBON

ENERGY: GLOBAL PROSPECTS 1985-2000. Workshop on Alternative Energy Strategies. McGraw-Hill Book Company, 1977.

ENERGY SUPPLY TO THE YEAR 2000: GLOBAL AND NATIONAL STUDIES. Workshop on Alternative Energy Strategies. The MIT Press, 1977.

SURVEY OF ENERGY RESOURCES 1976. World Energy Conference. United States National Committee of the World Energy Conference, 1977.

LA FORMACION DE MEMBRANAS CELULARES

MEMBRANE ASSEMBLY: SYNTHESIS AND INTRACELLULAR PROCESSING OF THE VESICULAR STOMATITIS VIRAL GLYCOPROTEIN. F. N. Katz, J. E. Rothman, D. M. Knipe y H. F. Lodish en *Journal of Supramolecular Structure,* vol. 7, págs. 353-370; 1977.

RAPID TRANSMEMBRANE MOVEMENT OF NEWLY SYNTHESIZED PHOSPHOLIPIDS DURING MEMBRANE ASSEMBLY. James E. Rothman y Eugene P. Kennedy en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 74, n.º 5, págs. 1821-1825; 1977.

MEMBRANE ASYMMETRY. James E. Rothman y John Lenard en *Science,* vol. 195, págs. 743-753; 25 de febrero de 1977.

SYNCHRONIZED TRANSMEMBRANE INSERTION AND GLYCOSYLATION OF A NASCENT MEMBRANE PROTEIN. James E. Rothman y Harvey F. Lodish en *Nature,* vol. 269, págs. 775-780; 27 de octubre de 1977.

LA OPTICA DE ONDAS GUIADAS

INTEGRATED OPTICS. P. K. Tien en *Scientific American,* vol. 230, n.º 4, págs. 28-35; abril, 1974.

SPECIAL ISSUE: LIGHTWAVE COMMUNICATIONS en *Physics Today,* vol. 29, n.º 5, págs. 23-60; mayo, 1976.

COMUNICACIONES OPTICAS. W. S. Boyle en *Investigación y Ciencia,* n.º 13, págs. 24-33; octubre, 1977.

LA ESTRUCTURA PROFUNDA DE LOS CONTINENTES

THE CONTINENTAL TECTOSPHERE. Thomas H. Jordan en *Reviews of Geophysics and Space Physics,* vol. 13, n.º 3, págs. 1-12; agosto, 1975.

LATERAL HETEROGENEITY OF THE UPPER MANTLE DETERMINED FROM THE TRAVEL TIMES OF MULTIPLE *ScS*. Stuart A. Sipkin y Thomas H. Jordan en *Journal of Geophysical Research,* vol. 81, n.º 35, págs. 6307-6320; 10 de diciembre de 1976.

COMPOSITION AND DEVELOPMENT OF THE CONTINENTAL TECTOSPHERE. Thomas H. Jordan en *Nature,* vol. 274, págs. 544-548; agosto 1978.

LAS TRAYECTORIAS DE LOS MOVIMIENTOS BRUSCOS DEL OJO

DERIVATION OF A MODEL FOR THE HUMAN EYE-POSITIONING MECHANISM. Gerald Cook y Lawrence Stark en *Bulletin of Mathematical Biophysics,* vol. 29, págs. 153-174; 1967.

NEUROLOGICAL CONTROL SYSTEMS: STUDIES IN BIOENGINEERING. Lawrence Stark. Plenum Press, 1968.

DYNAMIC OVERSHOOT IN SACCADIC EYE MOVEMENTS IS CAUSED BY NEUROLOGICAL CONTROL SIGNAL REVERSALS. A. Terry Bahill, Michael R. Clark y Lawrence Stark en *Experimental Neurology,* vol. 48, págs. 107-122; 1975.

TIME OPTIMAL BEHAVIOR OF HUMAN SACCADIC EYE MOVEMENT. Michael Clark y Lawrence Stark en *IEEE Transactions on Automatic Control,* AC-20, págs. 345-348; 1975.

GLISSADIC OVERSHOOTS ARE DUE TO PULSE WIDTH ERRORS. A. Terry Bahill, Frederick K. Hsu y Lawrence Stark en *Archives of Neurology,* vol. 35, págs. 138-142; marzo, 1978.

VIBRACIONES ACOPLADAS DE LAS CUERDAS DEL PIANO

DECAY RATES OF PIANO TONES. Daniel W. Martin en *Journal of the Acoustical Society of America,* vol. 19, 4, págs. 535-541; julio, 1947.

TUNING PREFERENCES FOR PIANO UNISON GROUPS. Roger E. Kirk en *Journal of the Acoustical Society of America,* vol. 31, n.º 12, págs. 1644-1648; diciembre, 1959.

FUNDAMENTALS OF MUSICAL ACOUSTICS. Arthur H. Benade. Oxford University Press, 1976.

COUPLED PIANO STRINGS. Gabriel Weinreich en *Journal of the Marine Biociety of America,* vol. 62, n.º 6, págs. 1474-1484; diciembre, 1977.

LA CABEZA DEL CACHALOTE

PHYSICAL PROPERTIES OF SPERMACETI OIL IN THE SPERM WHALE. M. R. Clarke en *Journal of the Marine Biolocal Association of the United Kingdom,* vol. 58, n.º 1, págs. 19-26; febrero, 1976.

STRUCTURE AND PROPORTIONS OF THE SPERMACETI ORGAN IN THE SPERM WHALE. M. R. Clarke en *Journal of the Marine Biological Association of the United Kingdom,* vol. 58, n.º 1, págs. 1-18; febrero, 1976.

TERMODINAMICA DE LA EVOLUCION BIOLOGICA

SELFORGANIZATION IN NON-EQUILIBRIUM SYSTEMS. G. Nicolis & L. Prigogine, John Wiley and Sons, New York, (1977).

A COOPERATIVE EFFECT WITH SIGNIFICANCE IN BIOLOGICAL EVOLUTION. J. Wagensberg en *Lectures Notes in Physics,* vol. 84, p. 350. Springer-Verlag. New York, 1978.

ENTROPY BALANCE IN BIOLOGICAL DEVELOPMENT AND HEAT DISSIPATION IN EMBRYOGENESIS. D. Lurié y J. Wagensberg en *Journal of Non-Equilibrium Thermodynamics* 4, 120; 1979.

NON-EQUILIBRIUM THERMODYNAMICS AND BIOLOGICAL GROWTH AND DEVELOPMENT. D. Lurié y J. Wagensberg en *Journal of Theoretical Biology* (en prensa).

JUEGOS MATEMATICOS

LOXODROMIC SEQUENCES OF TANGENT SPHERES. H. S. M. Coxeter en *Aequationes Mathematicae,* vol. 1. n.º 1/2, págs. 104-121; 1968.

EXCURSIONS IN GEOMETRY. C. Ogilvy. Oxford University Press, 1969.

TALLER Y LABORATORIO

PHOTOGRAPHY THROUGH THE MICROSCOPE. Eastman Kodak Company, 1974.

ADVENTURES WITH A MICROSCOPE. Richard Headstrom. Dover Publications, Inc., 1977.

Seguiremos explorando los campos del conocimiento



LA VERIFICACION DE LOS ACUERDOS SALT II, por Les Aspin
Los Estados Unidos disponen de un amplio conjunto de "recursos técnicos nacionales" de vigilancia, capaces de descubrir cualquier posible intento de la URSS encaminado a conseguir una superioridad significativa, violando los nuevos acuerdos.

ESTUDIO DE ESTRUCTURAS ATOMICAS DE SUPERFICIES CRISTALINAS MEDIANTE DIFRACCION DE ELECTRONES, por P.M. Echenique y M.A. Van Hove
Un sólido se comunica con el medio exterior a través de sus primeras capas de átomos: "su superficie". Dichas capas representan un papel crucial en muchos procesos físicos y químicos.

ARQUEOLOGIA DEL FONDO DE LOS LAGOS, por Aimé Bocquet
Los lagos de montaña, desde Francia hasta Austria, y desde Alemania hasta Italia, atrajeron a muchos pobladores prehistóricos. Los arqueólogos están ahora investigando lugares a los que sólo habían llegado buceadores coleccionistas.

SILVICULTURA, por Stephen H. Spurr
La gestión de un bosque mediante técnicas tales como la recolección selectiva de árboles, las entresacas prudenciales y la reforestación planificada puede duplicar y hasta triplicar la productividad del bosque.

SOLITONES, por Claudio Rebbi
Son ondas que no se dispersan ni se disipan, sino que mantienen su forma y tamaño indefinidamente. Se acaba de descubrir que los solitones pueden considerarse como partículas elementales de masa no nula.

EL ORIGEN DEL CAMPO MAGNETICO TERRESTRE por Charles R. Carrigan y David Gubbins
Se cree que es debido a una dinamo que funciona a base de un flujo de materia en el núcleo terrestre. ¿Qué es lo que impulsa dicho flujo?

LA INFLUENCIA DEL ESTETOSCOPIO EN EL DESARROLLO DE LA MEDICINA por Stanley Joel Reiser
El estetoscopio, primer instrumento utilizado de modo general para el diagnóstico, cambió la imagen de la enfermedad que el médico tenía y la relación de éste con el paciente.

COMO RESPIRAN LOS HUEVOS DE LAS AVES, por Hermann Rahn, Amos Ar y Charles V. Paganelli
Los embriones de las aves captan oxígeno y liberan dióxido de carbono por simple difusión a través de poros microscópicos de la cáscara del huevo.

INVESTIGACION Y
CIENCIA